

UNDER THE EDITORSHIP OF GARDNER MURPHY

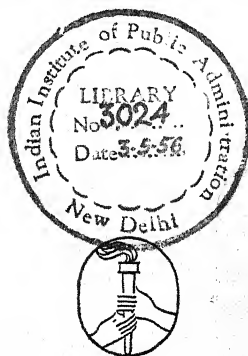
Principles of EMPLOYMENT PSYCHOLOGY

HAROLD ERNEST BURTT, PH.D

Professor of Psychology

Ohio State University

REVISED EDITION



IIPA LIBRARY



3024

HARPER & BROTHERS PUBLISHERS

New York and London

PRINCIPLES OF EMPLOYMENT PSYCHOLOGY, REVISED EDITION

Copyright, 1926, 1942, by Harold Ernest Burt
Printed in the United States of America

All rights in this book are reserved.

*No part of the book may be reproduced in any
manner whatsoever without written permission
except in the case of brief quotations embodied
in critical articles and reviews. For information
address Harper & Brothers*

B-Y

CONTENTS

PREFACE TO THE FIRST EDITION	vii
PREFACE TO THE REVISED EDITION	xi
I. INTRODUCTION	1
II. PSEUDO-PSYCHOLOGY	14
III. HISTORY OF SCIENTIFIC VOCATIONAL PSYCHOLOGY	52
IV. TYPES OF MENTAL TESTS	71
V. MENTAL TEST TECHNIQUE	122
VI. THE CRITERION	171
VII. THE SUBJECTS USED IN EVALUATING TESTS	200
VIII. SPECIAL CAPACITY TESTS: TOTAL MENTAL SITUATION	215
IX. SPECIAL CAPACITY TESTS: THE MENTAL COMPONENTS OF THE JOB	254
X. INTELLIGENCE AND VOCATIONAL APTITUDE	286
XI. INTERESTS IN EMPLOYMENT PSYCHOLOGY	318
XII. RATING SCALES	347
XIII. MISCELLANEOUS DETERMINANTS OF VOCATIONAL APTITUDE	405
XIV. TRADE TESTS	477
XV. JOB ANALYSIS	508
XVI. THE OUTLOOK FOR PSYCHOLOGY IN SELECTING PERSONNEL	535
APPENDIX I: Illustrating the Technique of Correlation	550
APPENDIX II: Illustrating the Derivation of a Regression Equation	559
INDEX	565

18
ms. Rom.

PREFACE TO THE FIRST EDITION

This book is an outgrowth, in the first instance, of material used for several years in presenting principles of employment psychology to college students, and in the second instance, of practical experience in personnel work and frequent contact with business men interested in psychology in so far as it relates to their problems. Effort is made, on the one hand, to give a fairly comprehensive account of the principles involved for the use of students preparing for practical psychological work in industry, and on the other hand, to avoid a discussion that is too technical for the reader without a psychological background. This does not mean, however, that the treatment is superficial. It is hoped, on the contrary, that the business man reading the book will realize the importance of a careful experimental approach to scientific employment psychology.

Statistical methods must of necessity form a part of the discussion. Although many persons shy at statistics, they are of such wide applicability in employment work that they cannot logically be omitted. No assumption of mathematical knowledge, however, is made, and the effort has been to make any statistical discussion as simple and as clear as possible. Wherever it has proved feasible to describe a method in a general way and relegate the more exacting details to the appendix, this has been done.

The critical psychological reader will notice that no definite stand has been taken regarding the fundamental points of view or metaphysical considerations of theoretical psychology. The author feels that these problems are not germane to the present discussion. The important thing is to predict occupational success whether this is construed from the standpoint of mind or muscle. It is pragmatically justifiable to speak of a "test of attention" regardless of the ultimate nature of attention or whether such a category exists at all. The employment psychologist's task is to

arrive at his practical goal regardless of the route taken. Most of us engaged in this field are too busy with our own problems to solve the fundamental issues with which other psychologists are better qualified to deal. In the following discussion it will probably be found that the methods are for the most part objective, but that the terminology is conventional. Experience in presenting psychological principles to the practical man has indicated the desirability of discussing them in terms of everyday vocabulary.

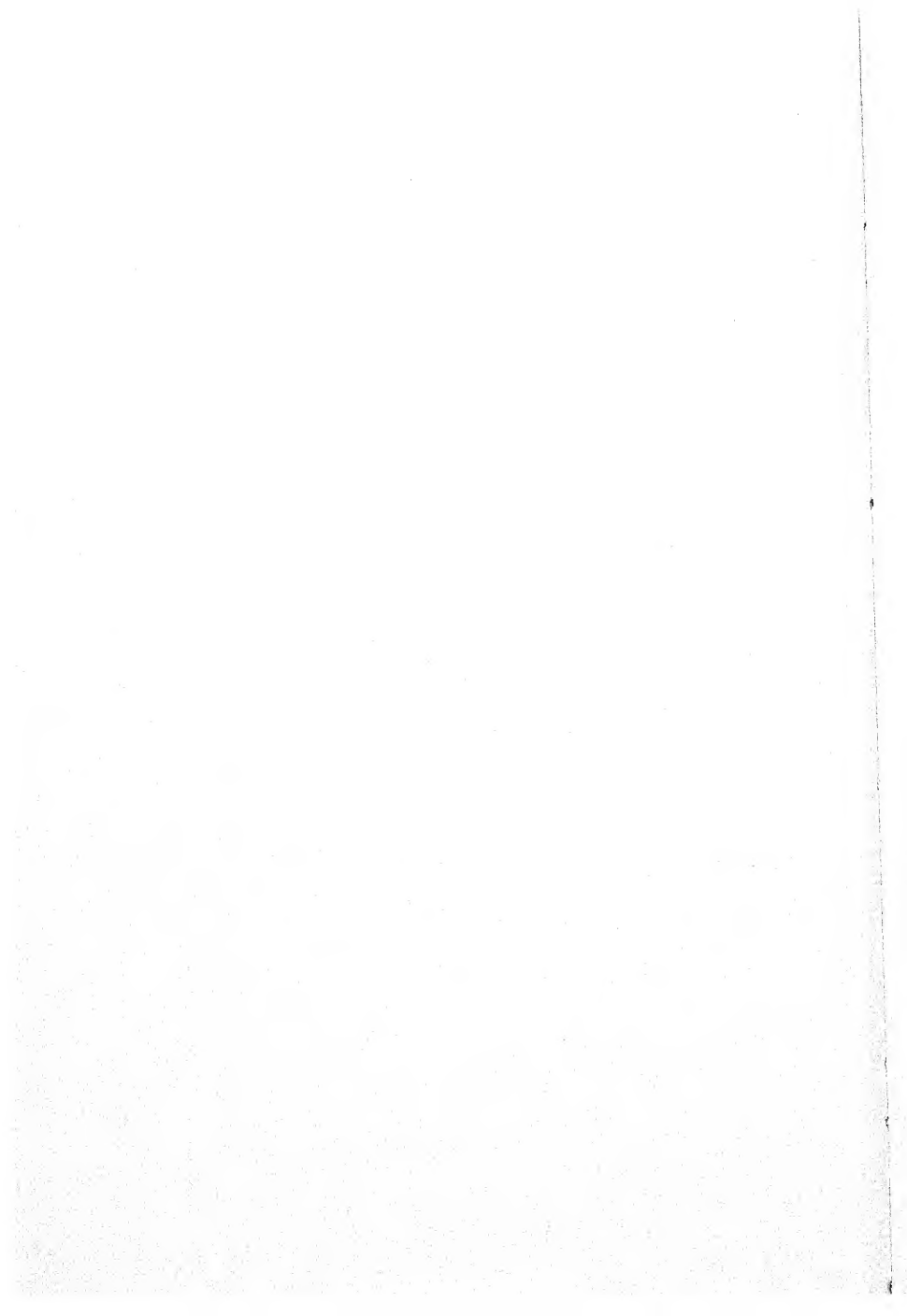
A work of this sort naturally draws rather heavily from experimental material contributed by many psychologists. Such studies as are referred to are cited mainly for illustrative purposes rather than in the nature of a critical review. To this end little mention is made of many details such as the time limit for a mental test, or the actual number of persons involved in a particular experimental study. This is done in order to avoid confusion from too many figures, and does not detract appreciably from the illustrative value of the citation. Reference is always made to the bibliography, however, so that the critical reader may, if he wishes, consult the original article and evaluate it for himself.

The author is indebted to all those who have contributed their bit to the body of psychological knowledge in this general field and whose results have been drawn upon rather extensively. He is especially indebted to H. L. Hollingworth, W. D. Scott, and A. J. Snow, whose contributions have been quoted more extensively. Grateful acknowledgment is likewise made to D. Appleton and Company, New York, for permission to quote from *Vocational Psychology*, by H. L. Hollingworth, *Judging Human Character*, by H. L. Hollingworth, and *Applied Psychology*, by H. L. Hollingworth and A. T. Poffenberger; to the McGraw-Hill Book Company, Inc., New York, for permission to quote from *The Selection and Training of Salesmen*, by H. G. Kenagy and C. S. Yoakum; to James P. Porter, editor, for permission to quote from *The Journal of Applied Psychology*; to A. W. Shaw Company, Chicago, for permission to quote from *Personnel Management*, by W. D. Scott and R. C. Clothier, and *Psychology in Business Relations*, by A. J. Snow; and to The Williams and Wilkins Company, Baltimore, for permission to quote from *Ability to Sell*, by M. J. Ream, and from *The Journal of Personnel Research*.

It is hoped that the book will show the practical man the importance of painstaking scientific technique in employment psychology in contrast with the expeditious but unreliable methods of unscientific pseudo-psychology. On the other hand, it is hoped that students who expect to pursue psychology in a practical way will find herein a fairly adequate background for plunging further into details.

HAROLD E. BURTT

Columbus, Ohio



PREFACE TO THE REVISED EDITION

There has been much activity in the field of personnel psychology since the first edition of this book. A little of this activity has involved the development of new principles such as factor analysis. Most of the work, however, has consisted of more extensive use of existing principles. The methods have been applied to a much wider range of occupations and many new tests or other predictors have been developed and validated. In the field of personality measurement there has been quite a bit of progress. Considerable impetus to the whole program has resulted from cooperative personnel ventures such as some of the work of the U.S. Employment Service. Various government agencies to an increasing degree have utilized psychological personnel procedures.

With the general principles remaining much the same as at the time of the first edition, the present revision involves no radical change in topics covered. Thus the same chapters appear, although they have been completely rewritten. Methods that are no longer in use have been dropped. Such new ones as have been developed are discussed. More recent illustrative material is incorporated.

The original bibliography was fairly complete in 1926. A complete bibliography at the present time would be almost prohibitive and scarcely worth while. Many of the early references would be useless to anyone now and have been dropped. From the more recent ones, careful selection has been made of those which are most pertinent to the present discussion. For the most part, citations are included only for articles and books to which reference is made in the text. With this objective it seems better to include the references at the end of each chapter near the point where they are apt to be used rather than in a single bibliography at the end of the book.

The only direct quotations in the revised edition not covered by prior acknowledgment are from the *Journal of Applied Psychology* and the *Personnel Journal*. Grateful acknowledgment is hereby made to the editors of these journals, James P. Porter and Charles S. Slocombe respectively, for permission to use a considerable amount of material.

HAROLD E. BURTT

Columbus, Ohio

PRINCIPLES OF EMPLOYMENT PSYCHOLOGY

Chapter I

INTRODUCTION

General Psychology. Psychology studies human experience and behavior. It endeavors to describe facts and to derive general laws for predicting how one will feel, think, and act under given conditions with a view to controlling those feelings, thoughts, and actions by controlling the conditions. From the time of classical Greece until the nineteenth century it was primarily an adjunct of philosophy and its method was *speculative* and casual. The science has now, however, moved from the armchair to the laboratory and is distinctly *experimental* in character. If the early psychologist was interested, for instance, in the bodily accompaniments of emotion, he sat and imagined himself in some dangerous or otherwise emotional situation and tried to observe his bodily feelings. The modern psychologist approaches the same problem by recording on a moving tape the pulse, breathing, blood pressure, and involuntary movements of the person on whom the experiment is being conducted, and then induces emotional states by moving pictures, snakes, revolver shots, or by providing a situation in which the person must sometimes lie and sometimes tell the truth. The early psychologist investigated color vision by looking at the sunset. The modern psychologist throws a beam of light through a prism and with narrow slits selects from the resulting spectrum bands of colored light of known wave length, varying their energy to determine the effect on visibility. To study the process of association, the early psychologist looked at some object such as a tree and noted what ideas came to him as a result. The modern psychologist uses apparatus which suddenly exposes a typewritten stimulus word and measures in thousandths of a second the time between the instant of exposure and the instant an observer speaks into a

diaphragm the first associated word that comes to him. The early psychologist was content with a few casual observations. The modern psychologist often makes hundreds and subjects the results to rigorous mathematical treatment. In almost any psychological laboratory are to be found precision instruments, adaptations of electrical and mechanical principles to specific problems, printed blanks for standardized mental tests, and statistical equipment.

Applied Psychology. The advent of applied psychology is more recent. Almost no practical use was made of psychological principles until the present century. There are several reasons for this. In the first place, there could be no application of the science until there were some principles to apply. A certain theoretical background is necessary in any science before it reaches the practical stage. Alchemy and necromancy are instances of premature efforts to apply a science. It must be remembered that, although there was some experimentation prior to that time, the first actual psychological laboratory was established in 1879. There were many psychologists as late as 1917 who believed that the theoretical basis had not been sufficiently laid for an applied science, and were loath to consider such a thing as military psychology.

A second factor that delayed the advent of applied psychology was the charlatan. Many a worthless proposition for improving efficiency, analyzing character, or curing ailments was presented under the guise of psychology. People invested in such propositions and were subsequently disappointed. Consequently, when a real applied psychologist approached them they recalled their earlier experience with "psychology" and failed to react favorably to his proposals. Pseudo-psychology will be discussed in more detail in the next chapter. The point in the present connection is that these "gold bricks" injured the reputation of the real psychologist and made it difficult for him to make progress in his practical contacts.

A third factor involved in the late development of applied psychology was the emphasis on general laws rather than on individual differences. Following the lead of the other sciences, the general principles were studied first. It was quite natural that interest should first center, for instance, on the general relation between memory and the method by which a poem was studied—

as a whole or piecemeal—rather than on the fact that one individual possessed different memory ability from another. It was likewise to be expected that the earlier experiments would be more concerned with determining to which sort of signal one could react more quickly—auditory or visual—rather than with ascertaining whether one individual could react a few hundredths of a second more rapidly than another person. Yet it is these latter aspects that are often of greatest interest to the applied psychologist. He is concerned with such things as the intelligence of an individual child who is backward in school, the early emotional experience of a particular patient with an obsession, the changes in blood pressure of a given criminal suspect during examination, or the attention and reaction time of a certain prospective employee. Until there was a partial shift from the study of general laws to the investigation of individual differences, the time was not ripe for applied psychology. Recent years, however, have witnessed marked advances in the contact of psychology with education, law, medicine, and business.

Psychology in Industry. Modern industry is especially concerned with three things—raw materials, equipment to construct the product from these materials, and human beings to operate the equipment, and buy and sell the finished product. The first of these involves such sciences as geology, botany, chemistry, and economics; the second falls especially within the sphere of engineering; but in the third there has developed of late years a realization of the importance of psychology. This importance extends in three directions: (1) selection of personnel; (2) industrial efficiency; (3) advertising and selling. The first of these involves primarily the placement of persons in the type of work to which they are best adapted. The second involves giving the person thus placed a chance to realize his maximum efficiency by proper adjustment of the methods and conditions of work. It involves such problems as the training of workmen, economy of movement, reduction of fatigue and monotony, the effect of ventilation or illumination upon efficiency, the maintenance of morale, and the reduction of accidents. The third involves controlling the prospect's attention, making him remember the proposition, and discovering basic motives and desires that will lead him to purchase a specific product. The present book is confined

entirely to the first of these three aspects of industrial psychology—the selection of personnel.

Psychology and Personnel Selection. The need for psychology in selecting personnel is obvious. Every employment manager and every foreman is familiar with the occupational misfit—the square peg in the round hole. The explanation of the presence of such misfits in industry is simple. Different jobs require for their satisfactory performance different mental and motor capacities. Individuals differ in mental and motor capacity, and it is frequently the case that the capacity necessary for the job and the capacity possessed by the person working at that job do not correspond. Suppose, to take an oversimplified example, that good memory is absolutely necessary for success in a given job and that applicants with good memory and with poor memory are available in about equal numbers. If they are hired at random, about half of them are doomed to failure because they lack the requisite memory ability. A careful survey of almost any large plant would reveal many a workman with slow reaction time vainly trying to keep up with a rapidly operating machine, or a man with defective attention attempting to concentrate on a task that is too complex for him, or a person with intelligence inadequate to grasp the problems and make the decisions necessary in his work.

The remedy obviously consists in placing a man in a job requiring aptitudes which he possesses. The management may know pretty well the requirements of the job just as it knows the requirements of the raw materials, but while it measures the tensile strength of the fabric and the specific gravity of the compound, it makes no effort to measure the mind of the workman who is handling that fabric or compound. There was good reason for this a few years ago because methods of mental measurement were not available, but this is no longer true. The development of mental tests, rating scales, and statistical techniques has opened up a wide field for scientific contribution to the problems of employment.

The Problem of Employment Psychology. Determining what mental capacities are needed for a given occupation and devising methods of measuring those capacities constitutes the problem of employment psychology. These measurements may

then be used upon applicants to determine their probable success in the occupation. Instead of hiring a man without consideration of his mental qualifications and waiting for time to show whether or not he is wisely placed, it is possible at the time of hiring to make some prediction of his ultimate success in a given job. The bulk of our present industrial work does not involve actual trade skill, but rather a limited number of operations which the worker must learn after he is hired and the performance of which depends largely upon his innate *capacity* or *aptitude* rather than upon any *proficiency* he has acquired in school or in a previous occupation. It is this type of personnel problem which is most frequently approached by the psychologist. The technique of mental tests or measurements of innate capacity is widely used in this connection. There is a further type of personnel problem which necessitates trade tests. This need arises when selecting workers such as carpenters or machinists who, at the time they are hired, profess some trade proficiency. It is desirable to determine by a trade test whether they actually have the proficiency which they claim.

Fundamental Principle of Employment Psychology. There is one principle that is fundamental in dealing with the above problems. The tests or other measurements to be used in selecting persons for a given occupation must be evaluated by giving them to persons whose actual ability in that occupation is known and *comparing efficiency in the test with efficiency in the occupation*. In other words, we must not devise a test that seems plausible, trust that it will work, and start using it for employment purposes. We must first *test the test*. If workmen who are good in the test are good in the occupation and those who are poor in the test are poor in the occupation, then the test is valid, while if there is no consistent relation between occupational ability and test score the test is useless. In the latter case the test is scrapped. In the former case, if the test is given to a prospective employee who has never worked at the occupation in question and he makes a high score, it is fairly safe to predict that he will be successful in the occupation after he has learned it, while if he makes a low test score it is probable that he will be unsuccessful even after long training. The procedure is, of course, not as simple as outlined here; subsequent chapters will dis-

cuss the methods in considerable detail. However, this principle of testing the tests is central to the whole problem and its observance marks the difference between a scientific and an unscientific psychological approach to personnel problems.

Where Employment Psychology Is Most Valuable. If an establishment is contemplating the introduction of psychological methods of employment, the question naturally arises as to where it will be most profitable to begin. It is not feasible and probably not worth while to devise psychological methods for every job. It is naturally desirable to place the effort where it will do the most good. This involves two problems: (1) determining where the need is greatest and (2) determining where conditions are such that psychological methods will be valid.

The management usually has a pretty good notion of the locus of the greatest need. A high labor turnover often indicates occupational misfits. Other things, of course, contribute to turnover, but the square peg in the round hole is no mean factor, and it is usually possible to determine whether the other factors are important in a given case. The need for these employment methods depends further on the relation between applicants and vacancies. If the number of applicants for work of a given sort is no greater than the number of vacancies, selective methods are unnecessary because no selection can be made. Everyone who applies must be hired. If, however, the applicants exceed the vacancies in number, it is necessary to hire some and reject others. There is then opportunity for psychological methods to aid in the selection of those who have the greatest promise of success.

From the standpoint of psychological technique two considerations are involved in determining where such methods will be valid. In the first place, the measurements must be standardized on a considerable number of workmen. There is a danger in statistics of basing results on too small a number of observations. A meteorologist would not measure the temperature for two days during one summer in order to predict the temperature the next summer. It would be equally absurd for a psychologist to standardize a mental test on two lathe operators, a good one and a poor one, with a view to predicting the ability of others who were tested. Theoretically one should use a sufficient number

of operators so that if any more were added to the group the results would not be significantly changed. There are statistical devices for correcting the results to take account of the small size of the sample, but it is better to have an adequate sample at the outset. Few psychologists would be content with less than twenty individuals; fifty are better. Consequently, if only a few persons are working at a given job it is inadvisable to try to standardize upon them any occupational tests for that job.

The other technical consideration involved is the attitude of the workers on whom the measurements are being standardized. A mental test is worthless unless the person taking it does his best. Tests are designed to measure a person's maximum capacity of a given sort. If one does not exert himself to maximum effort the results are meaningless, for the superiority of one person to another in test score may signify merely that he tried harder and not that he possessed any superior ability. Consequently, if the workers on whom the test is to be standardized are hostile so that they will intentionally do poorly or make no effort to follow directions, or if the proposition cannot be presented to them in such a way that they will take it seriously, it is better not to attempt it at all. This arousal of proper attitude in many cases, however, calls merely for tact on the part of the management and the psychologist. When workers understand the real purpose of the testing program, they will realize that it is being carried out for the advantage of prospective workers as well as of the management. They will appreciate it as a serious matter and will cooperate. Having thus determined where the need for more efficient selection of employees is greatest and where there is a good prospect of valid results, the employment psychologist may then embark upon his program of testing the tests.

Industrial Psychology and Human Welfare. But the program to be discussed has still wider implications in the social order, and before plunging into details it will be well to consider employment psychology from the broad standpoint of its contribution to human welfare. This is desirable because of the feeling existing in some circles that any methods aimed at increased industrial efficiency are one-sided—benefiting the employer but not the employee; and further, because of the prevalent impression that such methods treat the worker as a machine and evaluate

and dispose of him in automatic fashion. These notions are without foundation as far as the psychologist is concerned.

Such methods, to be sure, are usually initiated by the management, obviously because it expects them to work to its financial benefit. But this does not mean that the principles adopted will not also benefit the employee. Both employer and employee are tremendously concerned with proper placement of the individual. The occupational misfit is an economic loss both to the company and to himself. He naturally decreases production, but he also decreases his own pay. He has little prospect of advancement, and the ultimate outcome is often his dismissal or his voluntary separation from the concern. He might have spent his time more profitably in learning an occupation for which he was better adapted. It may sometimes seem to be an immediate hardship to refuse a man a job for which he is unqualified, but it is doubtless a kindness to him in the end. Moreover, it is frequently a question not of rejecting him altogether, but rather of finding some other place in the plant where he will qualify. Economic waste hits all of us including the workman himself, and there is no waste more far-reaching than misdirected human activity. Psychology tries to alleviate this misdirected activity by placing the individual in that particular occupation where he stands the greatest chance of success.

Vocational adjustment may proceed from either end. We may take the individual and attempt to determine in which one of many vocations he has the greatest promise of success. This is usually termed "vocational guidance." Or we may take a group of applicants for a job and determine those that are best qualified. This is usually called "vocational selection." The present book is concerned only with the latter. The two fields, however, are not unrelated. As vocational selection develops standards for hiring people for various jobs, those standards can be used subsequently in guiding individuals. If, for instance, tests or other methods have been devised for selecting machinists and salesmen, it will be possible to give both sets to a youngster seeking a vocational objective and tell him in which of these directions he stands the greater chance of success. It will be many years, of course, before occupational standards are developed in sufficient numbers to make possible a comprehensive vocational

guidance program along these lines, but the results can be used as rapidly as they become available. Progress in this direction will be facilitated as tests are developed which are unique, i.e., independent of each other, and which can then be weighted differently for predicting success in different vocations. There is another way in which the work of selection indirectly contributes to guidance. If a person is refused a job in which the prognosis is very unfavorable, this increases his chances of locating something for which he is better adapted, and if he is prevented from entering a number of vocations in which he would never have a future, he is more liable to land where he belongs. Employment psychology thus works to the interest of the employee as well as of the employer.

The other prevalent notion that the scientific employment process is automatic and mechanical in character should be somewhat tempered. To be sure, the techniques to be discussed in the following pages may give this impression to a certain degree. The procedure must in its large outline be rather objective and impersonal, but it is hoped that in many cases it will be supplemented by other factors and by the good judgment of the employment department. For instance, a worker who represents the third generation of the same family which has been employed in the mill constitutes a social factor that cannot be overlooked. An individual who is temporarily inefficient because of some disability should naturally receive special consideration. An applicant whose morale is temporarily disturbed by external factors should be treated as a special case. The notion of the square peg in the round hole is not to be construed literally as an absolute, inelastic proposition. To some extent the man influences the job and the job influences the man, and there are many instances where the fit was originally slightly imperfect, but where minor changes produced a very effective result.

The idea of a "worker-in-his-work unit" involves the worker's capacity, interest, and opportunity. The most satisfactory results will come about through the interplay of these three. The worker needs certain minimum capacities in order to stand any chance of success in his job, but he also needs opportunity to develop those capacities and possibly others, and he needs such interest in the work as will enable those capacities to function ade-

quately. It is even probable that in some instances ability will conflict with interest. The latter should then be treated with respect and it should not be forgotten that the worker is an individual. If he feels that bricklaying is the one sort of work in all the world that interests him, this should receive some consideration in his final placement. This problem usually is more acute for the vocational counselor than for the psychologist in the employment office. Interests, however, are often due to an individual's experience rather than to any innate factor and to that extent are perhaps somewhat less of a fixed entity than are his capacities. A person sometimes likes a given job because he started there originally or because he has friends who induced him to go into that particular work. Or perhaps he prefers to work in New York near the bright lights, although there are better openings elsewhere. If an applicant is manifestly unfitted for a given job but is tremendously interested in it, this situation calls for tact on the part of the employer in showing him his small chance of advancement in this line and his better possibilities in some other line. Effort may well be made in such cases to interest the applicant in another kind of work for which he has the requisite ability. Even if he starts out with a definite interest in a given job but without the ability, it is quite probable that in the course of time, when success has not come, his interest will wane. A common type of interest that leads to considerable confusion in vocational adjustments is the desire possessed by a great many workers for a white-collar job. In some circles there appears to be a certain social stigma attached to an occupation which involves a dark shirt. This stigma is entirely unfounded. The work of the man in overalls is often more of a social contribution than that of the man in the white collar. After all, an individual's greatest contribution is to be made in the line for which he is best fitted. It is better for a man to be an expert machinist than a poor lawyer or to be an efficient carpenter than an ineffective physician.

These problems of vocational adjustment have still wider implications in the social order and extend beyond the machine shop or the stitching room. The maladjusted worker constitutes a serious social problem. He is apt to be in economic difficulty and even in straitened circumstances because, if he is engaged

in work for which he is not qualified, he is likely to be penalized in his compensation. It is probable that this factor contributes materially to poverty and the ills that go with it. It may likewise contribute to even worse things. Many delinquents or criminals can be accounted for through economic failure. Typical of this class is the individual (often of low intelligence) who is hired for one job after another, but fails in each after rather extensive trial, finally becomes discouraged, and either shoulders a tin can and starts up the track or else begins with petty larceny and goes from bad to worse. Being refused at the outset some of these impossible jobs rather than being permitted to waste time trying to master them might have sifted the individual until he reached a place where he could fit. Furthermore, these maladjustments lead to dissatisfaction and unhappiness. A considerable portion of our industrial unrest is due to the fact that workers are not engaged in those types of work for which they are suited. The continuous uphill effort and the subtle feeling of not getting along, while old age and sickness and unforeseen emergencies stand in the offing, give the worker's life an emotional undercurrent that is undesirable. It may express itself in his attitude toward his family or toward his employer or toward his fellow man in general. Having a job for which he is adapted will appreciably alter this undercurrent of dissatisfaction and unhappiness.

So the employment psychologist is confronted with the immediate problem of selecting men for a particular job, but is also indirectly concerned with the more remote but more far-reaching social problem of vocational adjustment. If every factory operative, every office worker, every salesman on the road, and everyone at an executive's desk could be doing the type of work for which he was best adapted and in which he was most interested, the world would be a better place. The following pages will discuss psychology's modest contribution to these ends.

Outline. The next chapter discusses pseudo-psychology. It is desirable to dispose of these abortive efforts before proceeding to discuss scientific methods. There is so much misuse of the term "psychology" and there are so many things on the market purporting to be applied psychology that it seems best to clear the ground at the beginning. Chapter III sketches the history of

scientific vocational psychology. Inasmuch as mental tests play a large role in employment psychology they are dealt with in a general way before proceeding to their actual application. Chapter IV describes typical mental tests with which a psychologist should be familiar before engaging in employment research. Chapter V deals with general test technique—the devising and administering of tests.

As suggested above, the tests for a given occupation must be evaluated by comparing test scores with ability in the occupation. This latter factor is technically called the criterion. The methods of obtaining this criterion—estimates by foremen or production figures—and means for combining various criteria into a single one are discussed in Chapter VI. The next chapter considers the “subjects” or workmen on whom the measurements are to be standardized. The distinction between measurements of capacity and proficiency has already been made. The first of these may be divided into special capacity such as memory, attention, judgment or reaction time, and general capacity or intelligence. Chapters VIII and IX deal with special mental capacities in relation to actual vocational performance. The former discusses the case in which the attempt is made to devise a special test that reproduces the total mental situation involved in the job, and the latter the method of dividing and analyzing the job into its mental components and measuring them separately. The technique of comparing or correlating test and criterion is described and illustrated for various occupations. Chapter X treats general mental capacity or intelligence in somewhat similar fashion. A separate chapter (XI) is devoted to interests. The employment psychologist is beginning to realize that other things besides ability are of importance, particularly a person’s interest in and attitude toward his occupation.

There are many aspects of personality that we are at present unable to measure—such things as honesty, tact, leadership. Some information regarding them is often desirable in employment problems. It is at present necessary to depend in such cases on the judgments or estimates of persons who know the applicant in question. However, it is possible to obtain these estimates in fairly scientific fashion and to make them considerably more reliable than if obtained in the ordinary manner. These methods

are discussed in Chapter XII on rating scales. There are various miscellaneous factors somewhat related to vocational aptitude that are sometimes used in lieu of, or as a supplement to, tests. Some of these are considered in Chapter XIII—academic record, personal history blank, letter of application, recommendations, and the interview.

Methods of measuring proficiency as contrasted with capacity, i.e., trade tests, are discussed in Chapter XIV. The technique for devising and standardizing them is given, with examples of the different kinds. Chapter XV deals with job analysis and specifications. It does not cover the entire field and method of job analysis, but confines itself mainly to the place of psychology in the more complete program of job analysis. The last chapter deals with the present status and the future possibilities of psychology in the selection of personnel.

Chapter II

PSEUDO-PSYCHOLOGY

THE INTELLECTUAL UNDERWORLD

Mr. Barnum's famous estimate that "a sucker is born every minute" was conservative. Man has always been interested in the inscrutable and especially in the future. Prophets, soothsayers, oracles, and astrologers have flourished since the dawn of history. While some of them doubtless have been sincere, others have deliberately applied pseudo-scientific methods for their own aggrandizement or pecuniary advantage. Even today if a wealthy person became interested in psychological methods for analyzing the mental traits of his children or his employees and this interest received due publicity, he would shortly be waited upon by a delegation from the intellectual underworld. Some would propose to read the horoscope of the people in question, others to study the lines on the palm of the hand, others to feel the bumps on the heads under consideration; another would bring along a neurotic friend who could go into a trance and communicate with some deceased relative in the spirit world to see what *he* thought about it, while still others would present methods for predicting character and future success from the shape of the forehead, ears, nose, or chin, from bodily posture or gait or even from the tendency for the middle vest button to protrude or recede.

If our friend asked these various persons individually if their technique was psychological, they would undoubtedly answer in the affirmative. If he inquired whether their methods were infallible or whether they could predict with only a certain margin of error, they would assure him that error was a thing with which they were unfamiliar and for which they had no use. If he took the trouble to have a number of them make their observations

and predictions separately, he would probably find them contradicting one another on salient points. If, on the other hand, he chose at random and followed the advice of one member of the delegation in planning, for instance, his child's career or in promoting minor executives, he would doubtless find ultimate results markedly at variance with the prediction. It would be easy, then, to foresee his subsequent response to a person promoting a psychological service or marketing a scientific book on the subject. He had already "tried this psychology" to his own detriment. He could not be expected to discriminate between real psychology and pseudo-psychology because he had never studied the former and no scientist had ever enlightened him. That is why in the present discussion of psychology in relation to personnel it seems best to clear the ground at the outset—before telling what psychology can do, to tell what it does not do and call attention to a few psychology "gold bricks."

The Reason for Its Existence. The reason for the existence of this pseudo-psychology is obvious—it pays the one who is promoting it. It lends itself admirably to advertising and to commercial exploitation. The "prospect" is confronted with statements and proposals about his own mind. These intimate matters naturally arouse his interest and arrest his attention, and this is a first step in the sale. Furthermore, in the mind of the layman, a certain atmosphere of mystery surrounds "psychology." There is a natural credulity toward the unknown or little understood. This makes one somewhat prone to believe the indefinite statements of the "applied psychologist," and belief is a second important step in the sale. It has thus been possible to capitalize interest and credulity and lead persons to accept a proposition, disguised with pseudo-psychological terminology, which they would reject under other circumstances. Consequently, recent years have witnessed a mushroom crop of "applied psychologists" who never saw a laboratory or clinic but who pay big income taxes; an avalanche of literature about obtaining health, happiness, and success by the use of various "systems" of vibrations or mental dynamism; the development of institutions of learning which teach "divine metaphysics" and other subjects related to mental efficiency; and even the "mental broadcasting station" which broadcasts treatment or advice to subscribers.

Main Objection of Psychologists. The real psychologists object to this sort of thing primarily because it is presented under the guise of psychology. If it were called "galomalism" or given some other meaningless name and the gullible wished to invest, it would be nobody's business. But it is called "psychology," and when the promised improvement in memory fails to materialize, when the blonde employees fail to come up to expectations, when the position of the vest button proves to be a non-differential of the salesman's success, and when the inspirational phonograph records and the psychology hymns fail to raise the salary, then real psychology gets the blame.

The Extent of Pseudo-psychology. The present extent of this pseudo-science and its practitioners is serious. The writer occasionally crosses the trail of one of them. When addressing a civic luncheon club not long ago, one of the members told him after the meeting about an applied psychology club which had been running for a dozen years since its organization by a lady from Cleveland. They were still meeting periodically, having their exercises in concentration, and singing hymns or parodies about psychology such as one to the tune of *Onward, Christian Soldiers* which ended with the refrain:

To the Cause be faithful
Ne'er downhearted be,
One and all rejoicing
In Psy—chol—o—gy.

There are popular psychology magazines filled with encouragement and inspiration for those who are unfortunate or ambitious. Much advertising space is devoted to courses and systems. To a psychologist the advertisements are more interesting than the editorial material because they reflect the type of interest of the reader and indicate the sort of misstatements which can be made with impunity. In the larger cities one or more practitioners go through town every year, do some advertising, hold private consultations, give public lectures, and attempt to sell their services or their system. The writer once took advantage of a system that had an arrangement with a local store whereby a coupon secured at the time of purchase made it possible to obtain a character reading at a reduced price. The technique was based on a photo-

graph and a blank calling for brief personal information. The principal objective was vocational guidance. The following excerpts will suffice:

Your color indicates that you are enduring, passive, conservative and constant, friendly, submissive, vindictive, affectionate, and imitative. The following vocations are suitable for you: journalism, law, medicine, manufacturing, merchandising, teacher, office work, music and agriculture. . . . Your structure indicates that you are active, energetic, athletic, mechanical, constructive, industrious, and like to be outdoors. Choose your vocations from the following, selecting the one that fits you best according to your color: plumber, blacksmith, bricklayer, plasterer, contractor, carpenter, mechanic, salesman, explorer, builder and railroader. . . . Your physical structure being triangular indicates that you should be studious, artistic, literary, scientific. Being of such body structure you should endeavor to fit your color vocations in with the following: teacher, educator, scientist, artist, accountant, clerk, secretary and stenographer. . . . Don't rely too much on your mental ability in figuring, prove your conclusions on paper before making a statement.

Certain contradictions in the foregoing are obvious, as well as the generality of many of the statements. The tragedy of it is that some persons might receive a similar document and actually decide their whole destiny on such an irrational basis. The local Better Business Bureau took care of the matter in the present case. Scientists obviously should do more than have a good laugh at such projects because of their serious social implications.

It is beyond the scope of the present book to discuss all aspects of pseudo-psychology. Consideration will be given only to those which at some time or other have purported to contribute to vocational or employment problems. Critical discussion of methods for improving mental efficiency or of therapeutic devices must be omitted. Only one suggestion will be made which applies to these as well as to the vocational attempts of pseudo-psychology. If a person claims to be a "psychologist" it is comparatively simple to check the validity of such a claim. There are two national organizations of recognized psychologists, and practically anyone who is competent to render professional services in this field belongs to one or the other of them. The American Psychological Association was the original organization; it put a traditional

emphasis on academic psychology but is now sufficiently broadened to include all aspects. A more recent development is the American Association for Applied Psychology which overlaps the other organization extensively but is devoted to the professional aspects of the science. Both of these associations publish an annual directory; if an alleged psychologist is not listed in one of these directories, it would be in point to determine why. Few persons outside of these associations know enough psychology to be using it effectively in a practical way. Some member of the psychology department at almost any university or college has a directory and would be glad to inform anyone whether a particular name is listed therein.

ASTROLOGY

Astrology is one of the oldest pseudo-sciences that has been used in the effort to analyze character. It is still with us, however, and has actually appeared in the employment office. The writer knows a man who does most of the hiring for a large industrial concern and who has a theory that persons born in the spring are unsuited for certain disagreeable dusty jobs in the plant. His theory doubtless is based on some observation. The date of birth is listed on the personnel blank and it is possible that at the time of a termination interview it might catch the interviewer's attention. Inquiry revealed that this man had kept no record of persons who left on account of the dust and were not born in the spring. In fact, the writer was called in to get conclusive proof of the theory by interviewing one man who was leaving the job in question and admitted that he was born in April.

Books are available describing the procedures of character analysis by astrology. A typical guide includes the following for persons born in February: "They are very intuitive and good judges of character and human nature. They are successful in mercantile interests and enterprises. The best wives are born in this month, being always faithful and devoted. Great sincerity and power are possible for those born in this month. They will excel in music and art and should marry those born in October, January, or June." Equally inane horoscopes have been broadcast on the radio. The writer and a colleague from the Astronomy Department once burlesqued these broadcasts over a local sta-

tion, making the whole procedure what they considered utterly absurd and reading fictitious horoscopes which were supposedly humorous. Nevertheless, some people took it seriously; some listeners inquired, for instance, what it would cost to get the entire prediction for the next six months. One gentleman bared his soul regarding his domestic troubles, gave his own and his wife's birthday, and pleaded for help.

The astrologers have no scientific explanation of the connection between the date of birth and vocational aptitude and they do not take the trouble to check their predictions statistically. It would be possible to tabulate the birth dates of samples of persons successful in various vocations and compare the data with the claims of an astrology handbook. Until some such steps are taken there is no point in attaching any significance to vocational astrology. However, it might be well for a personnel man to determine whether any members of his department were actually using such techniques. Many a person, who in most respects is logical, gets an erroneous idea and pursues it enthusiastically without any scientific check.

SPIRITUALISM

Spiritualism plays a greater role in vocational and employment problems than is realized. Many people who should know better attend séances, seeking information and advice; and many of the questions asked fall within the personnel field. A late member of the New York Stock Exchange consulted a medium regularly before going to the Exchange or embarking on any important venture. Many localities have a practitioner who by spiritualism solves problems for large numbers of clients. While he does not work in the actual employment office his predictions may receive some consideration—for example, in problems of promotion or transfer at the executive level. Thus it is well to be familiar with the difficulties in this field in case one encounters a personnel man who attaches some significance to the contributions of a spiritualistic medium.

An occasional visit to a spiritualist meeting is enough to convince the scientifically-minded of the inanity of the whole affair. General suggestions are put forth and anyone who appears interested gets the message. The writer himself has secured messages

from the spirit world by helping the medium along. For example, in the trance the medium saw machinery about the head of someone in the rear of the room and inquired if anyone worked around machinery. The writer volunteered that he did, and then it was a matter of somebody named Briggs attempting to get through from the spirit world. The writer suggested that he did not know a Briggs but asked if it could by any chance be Brooks. After further listening the medium reported it was Brooks, whereupon some harmless message was received to the effect that he should be careful of the fingers of his left hand for the next two weeks while working around the machinery.

Futility Until Telepathy Can Be Demonstrated. While the psychical researchers have collected quantities of data purporting to be authentic communications from famous people who have died (including communications from Plato in English), one initial theoretical point must be considered. There is no presumption that a mind in this world can communicate with a mind in the spirit world until it can be proved that two minds in this world can communicate without some physical medium such as light or sound. Telepathy has been the subject of some experiment and more controversy. The earlier experiments yielded essentially negative results [3, 19].¹ In the usual experiment, one person observes certain stimuli such as cards with symbols upon them. This person is called the "agent." The other person is called the "percipient"; he tries to "read the mind" of the agent and state which symbol the agent is observing. Many variations of the procedure are possible, with many types of stimuli; but the crucial point is to determine the probability of guessing the correct answer and to compare the actual results with such probability. If, for example, it is a matter of the agent's taking a card at random from a deck and concentrating on whether it is red or black, the percipient to give the answer, we would expect the latter to *guess* correctly half the time. If, however, he answers correctly 75 per cent of the time, some other factor must be involved, possibly telepathy.

As mentioned above, the earlier results were essentially nega-

¹ Numbers in brackets indicate references at the end of the chapter. A number in italics indicates a particular page in the reference denoted by the preceding number.

tive. More recent experiments in one instance have yielded what are claimed to be positive results [16]. These studies have been reported quite extensively and have created considerable discussion. This has been promoted by non-psychological publications fully as much as by those in the scientific field. Agreement has by no means been reached. The results have been criticized on many grounds, particularly mathematical. There is also the complication that when going through a deck of cards, for example, and guessing which of four symbols appears on the successive cards, we cannot assume that each judgment is psychologically independent of the others. If one has had spades several times in succession, this may affect his tendency to guess spades on the next choice. It is difficult to devise a mathematical procedure to take account of this cross-influence of psychological judgments upon each other. Some of these difficulties are discussed in reviews of this work [17]. It is not worth while here to evaluate this extensive literature. The only point is that telepathy has by no means been accepted by the majority of the scientific psychologists. While the question may still be open, we certainly are not safe in saying that telepathy has been scientifically demonstrated. Until such time as it is demonstrated, there is no use in talking about spiritualism, which implies communication between two people when one of them is out of this world altogether. Persons who accept advice from the spirit world for vocational or other purposes are flying in the face of science and putting themselves at the mercy of ignorance or unscrupulousness in the form of a medium.

PHRENOLOGY

Phrenology is another type of pseudo-science that is still current. Not many years ago a concern in New England and another in Texas engaged a phrenologist to work in the employment office. The writer himself was on one occasion mistaken for a phrenologist. When it became noised about the office and factory that a psychologist was to begin work, a number of persons, it was discovered later, expected to have the contour of their skulls examined.

Semblance of Scientific Basis. Phrenology did have historically a little more semblance of a scientific basis than the other pseudo-

psychologies mentioned above. Science had discovered that certain parts of the brain were concerned with certain sensory or motor functions. If a portion of the skull was removed and the surface of the brain stimulated, movements of certain muscles might take place, and by stimulating different parts of the brain different muscle groups could be made to contract. Moreover, injury to a certain portion of the brain often left a person with some defect such as inability to see or hear or speak.

Now when real scientists were presented with these facts, they set out to analyze the matter further by experiments on the brains of living animals, by post-mortem examination of the brains of people who during life had some mental or motor defect, and by dissection and microscopic examination to trace neural pathways from the sense organs and muscles to their destination in the brain. It was slow work and it is not yet completed. But all the phrenologist needed was a good start afforded by the knowledge that there was at least some brain localization no matter of how coarse a variety. It seemed plausible enough that if there was a brain center for movement of the arms there should likewise be centers for memory, reverence, combativeness, conscientiousness, philopropriogenitiveness, etc. Scientific method was too slow and laborious for the phrenologist. He made casual observations of his acquaintances, noting a little cranial protuberance here and there and attempting to find some mental trait of the individual to correspond; but he neglected to ascertain whether any people with a similar protuberance lacked the trait or whether any with the trait lacked the protuberance. Thus he built up a system and mapped out the skull in an utterly illogical and unscientific fashion. This movement started about 1800 and there has been little revision of the principles originally laid down. A book written in 1832 is still the standard today!

Assumptions of Phrenology. At least three assumptions made by phrenology are erroneous. In the first place, it assumes that there are a great number of specific traits or faculties whose function is located in a particular portion of the brain. All the evidence of scientific experiment, however, shows that the brain does not function in as small units as those claimed. It has been possible to locate regions concerned with various muscle groups and with vision, hearing, and most of the other senses. But no

detailed areas have been found to be concerned with such things as high vs. low tones or sensations of red vs. blue. A map of the functional areas of the brain made by the scientist is simple compared with that made by the phrenologist. Moreover, there are some parts of the brain with which no very definite function has as yet been found to be correlated, but phrenology long ago mapped the entire surface. An idea of the discrepancy between

TABLE 1. SCIENTIFIC VS. PHRENOLOGICAL STATEMENTS AS TO FUNCTION OF CERTAIN REGIONS OF THE BRAIN

Actual Function as Determined by Experiment	Function Alleged by Phrenologists
Movement of feet and legs Movement of trunk and shoulders Movement of hand and fingers Movement of jaws and lips Auditory sensations Touch, temperature, and muscle sensations Visual sensations Maintenance of equilibrium	Reverence Marvelousness Ideality Constructiveness Destructiveness Hope Love of children Amativeness

the actual findings of science and the assumptions of phrenology may be obtained by a detailed consideration of a few regions of the brain. The first column in Table 1 lists the functions of certain brain regions that have actually been determined, and the second column gives the corresponding functions assigned to these regions by the phrenologist. Attention is called especially to the location of "reverence" in the region actually concerned with the movement of the feet and legs, and of "amativeness" in the region actually concerned with the maintenance of equilibrium. Furthermore, the phrenologist locates memory in the front lower central part of the brain, a region the function of which has not as yet been scientifically ascertained. It is actually found that the localization of memory follows that of the sense department involved— injury to the visual region of the brain causing disturbance of memory for visual details but not for auditory details.

The second erroneous assumption of phrenology is that there is a direct and obvious relation between the development of a

trait and the size of the corresponding region of the brain. There is, to be sure, evidence of a slight relation between the size of the brain and intelligence, but the complexity of structure is equally important. When it comes to the development of the small regions with which the phrenologist is concerned, the difference, if any, would be practically imperceptible. For instance, it is pretty well established that speech is controlled by an area on the left side of the brain. Microscopic work indicates that the layer of gray matter in the corresponding region of the right side is not quite as thick, but the difference in thickness is not over a millimeter. No phrenological methods could detect a difference of this size.

The final assumption is that a few casual observations afford a sufficient basis for generalization. This is, of course, contrary to all scientific method, which implies the collection and statistical treatment of large numbers of observations before drawing conclusions. Books on phrenology comprise an analysis of a relatively small number of individual cases rather than a statistical treatment of large numbers of persons. The absurdities to which this technique has led are manifest in Table 1. So while phrenology has a more plausible basis than the other pseudo-psychologies, its fundamental assumptions are unsound and it has absolutely no contribution to make to scientific employment methods.

PHYSIOGNOMY

Physiognomy is the most widely used of these questionable methods of analyzing character or predicting mental capacity. If it is construed in a wide sense to include the appearance of the face and head and entire body, its use will be found quite widespread. Many firms require a photograph with the application blank in instances where the person is not available for an interview, or use the photograph to select those who are to be interviewed. In some types of work attractive personal appearance is, of course, a requisite, or race may be significant; but there is often a feeling that something of further value may be obtained from observing the photograph. Probably some aspect of the features influences the judgment, consciously or otherwise, of the one evaluating the application. The head of a large technical school arranges his interviews with the boys who apply for entrance in such a way that they have to walk down a long aisle

before reaching his desk. He believes that he obtains valuable insight into their traits or capacities by observing their gait during their approach. One employment man has an antipathy to red hair. An office manager eschews blondes in his organization. It is, then, worth while to consider scientifically the value of such methods for employment or vocational purposes. More detailed discussion is warranted than in the case of the other pseudo-scientific methods.

Popular belief in physiognomy is doubtless back of its more practical use and commercial exploitation. In our literature and in our personal contacts we have been taught to attach significance to the shifty eye, the high forehead, the receding chin, the dimple, the heavy jaw, the short neck, and even the erect posture or the shambling gait. These beliefs have developed, like many of our other unscientific notions, as a result of casual observation combined with an absence of logic. When we come to consider character from the practical standpoint of employment, we merely carry over uncritically the notions that have developed in popular thinking. The basis of these popular beliefs may now be analyzed in a little more detail.

Association by Similarity. It is a fundamental law in psychology that one thing is apt to suggest or call to mind another which is similar to it. Thus the photograph of a friend suggests that friend because the contours of the former are similar to those of the latter. "Robin" may suggest "oriole" or "cat" may suggest "tiger," for the same reason. This principle of analogy or association by similarity operates in our popular notions about physiognomy. A person with a short neck suggests a bull and we then attribute to him some of the aggressive characteristics of that animal. Cats are crafty and treacherous and clams are cool, flabby, and inert; hence arises the importance which we attach to the feline tread or the clammy handshake. By the same principle of similarity a broad forehead suggests a broad mind, hard-textured flesh suggests a hard heart, and sharp features a sharp, penetrating intellect. Or again, if the physiognomy of a stranger is like that of an acquaintance it is quite natural to attribute to the former the traits of the latter. If one has had a disagreeable personal experience with some red-headed person, he may assume that another "titian" is likewise irascible. These popular generali-

zations, then, are readily explainable by the law of association, but this does not justify them. The fact that one thing reminds you of something else does not establish it as a scientific truth that there is any actual relation between the two. The popular mind, however, is content with its assumption; and if sometimes the relation proves subsequently to hold and sometimes the reverse, it is customary to remember the former instances and to forget the latter.

Observation Influenced by Expectation. Another principle which is involved in the development of popular physiognomic notions is that we tend to see what we expect to see. If our attention is set for some particular aspect of an object, it is that part which we see first or which impresses us most vividly. In a familiar laboratory experiment in which a pointer swings along a scale and a bell rings at some particular point, if an observer is attending to or thinking about the bell he will judge that it sounds at an earlier position of the pointer than he will otherwise. Attending to the bell facilitates its entrance into consciousness. Again, if one attends to the trombone in an orchestra he can hear it stand out from the other instruments. A motor mechanic will detect a main bearing knock that the layman would overlook because the mechanic takes an attitude of expectation. This principle operates to substantiate our beliefs in physiognomy. If a person shakes hands weakly we expect that he is going to show some vacillation, and while he perhaps manifests that trait no more than do other people with whom we come in contact we are "set" for it in his case and notice instances which would otherwise escape us. Or if we observe someone with large ears and have been taught that these denote parsimoniousness, we watch for instances which might be construed as manifesting that trait and magnify them, although our friends with small ears may be acting in a similar manner. But once we observe these expected traits they serve further to confirm our generalization, as another case "which proves it."

Evidence of Habitual Activity. There are some aspects of popular physiognomy, however, which seem to have an objective reason to account for them instead of being dependent purely on the association process or attention attitude of the person making or corroborating the generalization. It seems plausible at first

glance that certain habitual activities should leave their impression in observable form on the face or body. A studious person bending over his books for years may become round-shouldered. A pugilist may develop a tendency to look at his adversaries—and everyone else—with his head turned toward the left and bent slightly forward. A philosopher may contract his brows while he ponders until the wrinkle becomes permanent. A criminal may repeatedly avoid the gaze of his prospective victims till his eye becomes shifty. While it is perfectly true that certain habitual tendencies may affect the musculature permanently, a fallacy is involved when it comes to reversing the proposition and assuming that those with round shoulders are studious, that those with a sidewise gaze are belligerent, that those with wrinkled brows are philosophers, and that those with unsteady eyes are criminalistic. There are other things that might equally well cause round shoulders, such as crap-shooting; or that might produce an unusual position of the head, such as poliomyelitis; or that might wrinkle the forehead, such as nearsightedness; or that might cause the eye to shift, such as shell shock. Popular beliefs regarding physiognomy have then no scientific basis. They are used, however, by many persons in practical problems of predicting human characteristics and this uncritical use must obviously lead to many mistakes. Moreover, our popular notions pave the way for our acceptance of systems of character analysis that have been commercialized.

Commercial Systems of Physiognomy. It was quite natural that the astute purveyor of psychology gold bricks should avail himself (or herself) of the fertile field of physiognomy in which the seeds of popular belief were already sprouted. If persons had some notions regarding the relation between the face or figure and character, why not devise a detailed system—arbitrary, to be sure, and without scientific foundation—and sell it to them? This is precisely what was done. The promoters wrote books and articles, and gave lectures and, best of all, personal consultation and advice, using such criteria as the following:

Texture is a great classifier of humanity. The individual of fine hair, fine-textured skin, delicately chiseled features, slender, graceful body and limbs, as a general rule is refined, loves beauty and grace, and likes work either purely mental in nature or offering an opportunity

to handle fine delicate material and tools. On the other hand, the man with coarse hair, coarse-textured skin, and large, strongly formed features inclines as a general rule to occupations in which strength, vigor, virility, and ability to live and work in the midst of harsh, rough, and unbeautiful conditions are prime requirements. . . . Blondes as a general rule are changeable, variety-loving, optimistic and speculative, while brunettes are consistent, steady, dependable, serious, and conservative. . . . Poets, educators, and essayists will show a marked tendency to resemble the triangle in structure of head and body—both head and body wide above and narrower in the lower portions. Generals, pioneers, builders, engineers, explorers, athletes, automobile racers, aeronauts, and others who lead a life of great activity, will show a general tendency toward structure on the lines of the square—square face, square body, square hands. Judges, financiers, organizers, and commercial kings will show a general tendency toward structure upon the lines of the circle—round face, rounded body, and a tendency to roundness in the hands and limbs [7, 39]

A news magazine reports generalizations proposed by an alleged medical authority in Paris. A rectangular face denotes balance, firm will, courage, and masculinity. A triangular face with the upper portion broader and leading down to a point at the chin denotes an intellectual or a brainy person. A long, oval type of face indicates nervousness, and one shaped like a trapezoid with the upper base narrower than the lower suggests a calm temperament. Some of the proposals like the foregoing constitute a more literary restatement of popular beliefs and some of them are dogmatic assertions.

EXPERIMENTAL EVALUATION OF CHARACTER ANALYSIS FROM PHYSIOGNOMY

While the theoretical basis of such generalizations seems unsound and while the methods employ mere observations and not actual measurements of the physiognomic characteristics in question, the crucial point is to determine experimentally whether the alleged relations actually exist. Suppose that photographs are available of a group of intimate acquaintances who can give a pretty reliable estimate of one another. It is possible then to obtain an idea as to a person's status in each of a number of mental traits, that status being the combined judgment of his acquaintances. The photographs may next be submitted to judges

who have never seen the individuals in question and they may be required to estimate each person in each trait from his photograph. Then these estimates made from physiognomy may be compared with the actual traits as indicated by the combined judgment of acquaintances, to determine whether physiognomy under these conditions has any validity in indicating the mental traits.

This may best be done by the procedure of *correlation*. This is a statistical method for indicating the closeness with which any two variables or sets of traits or measurements are related. If, for instance, those who are rated by their acquaintances as most intelligent are likewise rated from the photographs as most intelligent, and vice versa, we speak of a *high positive correlation*. If those actually most intelligent are judged from the photographs to be least intelligent, and vice versa, we speak of a *high negative correlation*, while if there is no tendency one way or the other we speak of a *zero correlation* or *no correlation*. By the use of proper formulae² it is possible to compute a *correlation coefficient* (often expressed as *r*) which indicates not merely whether the correlation is high, low, or negative, but exactly how close is the relation between the two variables. A coefficient of 1.00 indicates perfect correlation—i.e., the person who is highest in one variable is correspondingly high in the other, the person who is next highest in one is proportionately high in the other, and so on down the list. From 1.00 the coefficient can range down through zero to -1.00, which indicates a perfect negative correlation.

Interpretation of correlation coefficients of different magnitude is a bit difficult for one unfamiliar with statistics. It is necessary to divorce oneself from any notion that coefficients represent percentages of something, that a correlation of .50 represents 50 per cent or is half as good as a correlation of 1.00. This is far from the case. An adequate interpretation involves probability theory, but a hint can be given here. Suppose two variables are corre-

² Appendix I illustrates the computation of such coefficients and also discusses the interpretation of correlations of different magnitudes. In the examples presented there, sets of scores in test and job are given. These are then ranked. In the present connection the original estimates on the basis of physiognomy or acquaintance consist of ranks so that the computation would begin with the third and fourth columns in the examples.

lated and it is desired to predict one from the other—for instance, from a worker's intelligence test score to predict his production in a job. We first write out an equation which expresses one variable as a function of the other. We can then substitute the intelligence test score in the equation and solve for the production score. If the two variables are highly correlated this prediction can be made with some validity, but a certain amount of error is to be expected. It is possible to determine the error of this sort involved in predicting production from the intelligence score when the correlation between them is of a specified magnitude. It is also possible to determine the error that would be made in prediction if there were no correlation between the two variables, that is, if we were merely guessing. We then compute the percentage by which it is possible, using the intelligence test, to reduce the predicting or forecasting error that would prevail if we merely guessed at production, or if the correlation was zero. (Cf. Example VI, Appendix I.) The forecasting efficiency for a few typical values of correlation coefficients is given in Table 2. The correlation of .50, for example, reduces the error of prediction 13 per cent over what it would be if there were no correlation, and a perfect correlation reduces it 100 per cent.

TABLE 2. FORECASTING EFFICIENCY OF VARIOUS DEGREES OF CORRELATION

Correlation Coefficient	Forecasting Efficiency (per cent)	Correlation Coefficient	Forecasting Efficiency (per cent)
.10	$\frac{1}{2}$.70	29
.20	2	.80	40
.30	5	.90	56
.40	8	.95	69
.50	13	.98	80
.60	20	1.00	100

In actual practice a coefficient less than .30 does not attract much attention, but when the coefficient is .50 the personnel psychologist begins to be interested. A further and more casual notion as to the meaning of correlation coefficients of different

magnitude may be obtained from the following consideration. Children of the same family resemble one another to some extent in physical characteristics. Twins resemble each other more strikingly in these respects. In some instances, measurements have been made of such characteristics and for brothers and sisters show correlations around .40. Similar procedures for twins yield correlations around .80.

Estimates of Miscellaneous Traits from Physiognomy. A few studies of the kind just mentioned may be cited. A group of 25 college women rated one another in a considerable number of fairly definite traits [6, 37]. Each individual took the names of the 24 others and considering, for instance, "neatness," selected the one she considered neatest of all and marked her 1, then selected the next neatest and marked her 2, etc., so that the 24 were arranged in rank order from the neatest to the least neat. Then the same thing was done for refinement, sociability, and a series of other traits, each one being rated separately and each woman ranking all the other 24 women. There were thus available for each woman 24 estimates of her possession of a trait; e.g., she had been assigned a ranking in neatness by all the other women. These 24 figures were then averaged to get the consensus of the entire group regarding that particular woman's neatness. Similar averages were found for her refinement, sociability, etc. This procedure was repeated for each woman. This combined judgment of 24 acquaintances might be taken as about the best statement of the real characteristics of the women that could be secured. These figures having been obtained, photographs of the 25 women of uniform style and size were submitted to a group of men who were totally unacquainted with the women involved. Each man ranked the individuals with reference to neatness as far as he could judge it from the photographs, marking the neatest 1, and the next neatest 2, etc. Then he ranked them with reference to refinement, with reference to sociability, etc., making his estimates entirely on the basis of the photographs inasmuch as he did not know the individuals at all. It was then possible to compare or correlate the ranks assigned by any one man on the basis of the photographs with the ranks assigned by the combined judgment of acquaintances. In exactly the same way the photographs were submitted to another group of women totally un-

acquainted with the original group and they ranked them as the men had done on the basis of the photographs.

To consider the matter first from the standpoint of the combined estimates of the judges rather than from that of the accuracy of the individual judge, the ranks assigned by all the men to any photograph for a given trait were averaged. These average estimates from photographs were correlated with the combined judgments of acquaintances above mentioned. The

TABLE 3. CORRELATION BETWEEN AVERAGE ESTIMATES OF TRAITS FROM PHOTOGRAPHS AND AVERAGE ESTIMATES OF THOSE SAME TRAITS MADE BY ACQUAINTANCES³

Traits	Estimates of Photographs	
	By 25 Men	By 25 Women
Neatness.....	.03	.07
Conceit.....	.10	.27
Sociability.....	.29	.29
Humor.....	.21	.45
Likability.....	.30	.45
Intelligence.....	.42	.61
Refinement.....	.50	.52
Beauty.....	.60	.49
Snobbishness.....	.58	.53
Vulgarity.....	.61	.69
Average.....	.36	.44

same was done with the women's estimates from photographs. The results are shown in Table 3. The first column lists the traits involved; the next column gives the results when the group of men are estimating the traits from photographs, and the last column gives the results when the group of women are using the photographs as a basis for judgment. For instance, the combined opinion of acquaintances regarding the neatness of the individuals in question correlates with the combined opinion of a group of

³ From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, Inc., New York.

men (based only on photographs) regarding the neatness of the same individuals to the extent of .03. The former combined opinion correlates likewise with the combined opinion of a group of women (based only on photographs) regarding the neatness of these same individuals to the extent of .07. Similar figures follow for the other traits.

It is obvious that estimates from the photographs are none too satisfactory for practical purposes. Moreover, the value of the estimate appears to depend on the trait. Vulgarity, snobbishness, and beauty seem to be estimated fairly well from the photograph, while quite the reverse is true of neatness, conceit, and sociability. One would hesitate to use physiognomic diagnosis of many of the traits indicated for employment purposes even if he could obtain 25 judges to make the physiognomic estimates.

Although the results are none too satisfactory when the estimates from photographs made by a group of 25 judges are pooled, the situation is much worse if we consider the validity of an individual judge's estimate. In the usual employment situation there are, at most, only a few persons who evaluate a given applicant from his physiognomy. Instead of using the average estimates from photographs as in Table 3, we may take the estimates made by one judge from the photographs with reference to neatness, for example, and correlate these estimates with the combined estimates of the acquaintances regarding neatness. To indicate the typical trend, 10 judges are taken at random and their individual correlations for three of the traits are given in Table 4. The estimates of intelligence made, for instance, by Judge A, using the photographs, correlate with the combined opinion of acquaintances regarding the intelligence of the same individuals to the extent of .51. The estimates of sociability made by Judge C correlate with the combined opinion of acquaintances regarding sociability to the extent of .05.

Inspection of the table shows a great variation between judges. Some of them estimate a trait from physiognomy fairly well and others rather poorly. For instance, Judge A's estimate of intelligence has a correlation of .51, while Judge D's actually has a negative correlation. In neatness the best judge is C with a correlation of .29; the worst is H, with a correlation of $-.09$. In sociability J is the best (.55) and I the worst (.00). Moreover,

TABLE 4. CORRELATION BETWEEN ESTIMATES OF TRAITS FROM PHOTOGRAPHS
MADE BY INDIVIDUAL JUDGES AND AVERAGE ESTIMATES OF
THOSE SAME TRAITS MADE BY ACQUAINTANCES⁴

Judge	Intelligence	Neatness	Sociability
A....	.51	.11	.39
B....	.11	.10	.08
C....	.15	.29	.05
D....	-.27	.06	.49
E....	.08	.24	.08
F....	.43	.41	.28
G....	.04	.11	.02
H....	.39	-.09	.32
I....	.22	-.08	.00
J....	.30	.02	.55
Average	.20	.12	.23

a judge who estimates one trait well may fail when another trait is involved. Judge A, for instance, is fairly competent to estimate intelligence (.51), but manifestly incompetent to estimate neatness (.11); J estimates sociability with some validity (.55), but his estimates of neatness have no validity (.02). Consequently, it would seem hazardous to attach much practical significance to physiognomic estimates of this sort made by one or at most a few individuals. All-round judges of character from physiognomy apparently are scarce.

Allusion was made earlier to the technical high school principal who made character analyses by observing the gait of the applicants as they walked down the aisle. A check on gait as related to one personality characteristic has been made [4]. By means of Maslow Personality Inventories 16 college women were selected on the basis of dominance—8 at each extreme. Personal interviews were also used in selecting these individuals from a group of 238. Motion pictures were then taken of each individual walking 50 yards. The films were shown to 99 persons who were required to judge whether the girl was dominant or non-dominant.

⁴From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, Inc., New York.

Part of the time the upper half of the body was obscured by a card during projection, part of the time the lower half was obscured, and part of the time the entire figure was visible. On the whole, dominance was judged only a little better than would have been the case if the subjects had merely guessed at it. The entire picture was a little better in this respect than either of the halves. Results certainly were not significant enough to be of any practical importance in personnel work.

Estimates of Intelligence from Physiognomy. It is possible to make a more careful check than in the foregoing instances with reference to estimates of intelligence from physiognomy because they can be compared with intelligence as objectively measured by tests, whereas physiognomic estimates of other traits such as neatness or sociability must be evaluated by comparison with judgments of acquaintances. In one such study [1], 63 managers, buyers, and assistants in a large department store were given an intelligence test somewhat similar to the Army Alpha (*infra*). Their photographs were then submitted to 12 graduate students interested in personnel problems. The student judges were required to estimate the intelligence of these business men from their photographs. After looking through the pictures to get a general idea, each judge selected the 7 most intelligent and the 7 least intelligent, then the 14 who were superior but not as good as the first 7, and likewise the 14 who were inferior but not as poor as the lowest 7. Arbitrary values were assigned to each of the four classes in order to handle the data statistically. The 12 estimates of the intelligence of a given manager were then averaged to get the combined opinion of the judges regarding his intelligence. A similar average was obtained for each of the other men concerned. These combined estimates of intelligence from photographs were then correlated with the actual intelligence as measured by the tests. The correlation coefficient was only .27. Moreover, it must be remembered that these results used only the extremes of intelligence and did not include the middle group. Had this been included, the correlation would probably have been smaller still.⁵ It would seem that even when a dozen persons pool their results, estimates of intelligence from physiognomy are almost worthless.

⁵ For statistical reasons beyond the scope of the present work.

The same experiment may be considered from the standpoint of the validity of the individual judge. One of the simplest methods is to note for each judge how many men he places on the right side of the average and how many on the wrong side; i.e., whether a man he rates in the best 7 or superior 14 is actually above the average in measured intelligence or not. These results are shown in Table 5. Judge A, for instance, places 27 individuals

TABLE 5. NUMBER OF PERSONS CORRECTLY PLACED AS ABOVE OR BELOW AVERAGE INTELLIGENCE BY INDIVIDUAL JUDGES ON THE BASIS OF PHOTOGRAPHS^b

Judge	Number of Photographs on Correct Side of Average	Number of Photographs on Wrong Side of Average
A....	27	15
B....	23	19
C....	25	17
D....	23	19
E....	19	21
F....	26	16
G....	22	20
H....	22	20
I....	20	22
J....	20	22
K....	22	20
L....	22	20
Total	271	231

correctly on the basis of these photographs; i.e., if they are actually above the average in tested intelligence he places them above on the basis of physiognomy, and vice versa. However, he misplaces 15 individuals, i.e., judges them as above average when they are actually below, or vice versa. There are only two or three judges who place many more of the men correctly than incorrectly and some actually have more in the incorrect column. The total for the correct column is only 17 per cent more than that for the incorrect column. So while pooled judgments are

^b After Anderson.

bad enough, individual judgments are worse, and it would manifestly be useless for one or two persons to use such physiognomic methods for practical purposes.

Another study of the relation between intelligence as estimated from photographs and as actually tested utilized 800 visitors at a national business show as judges of 10 photographs of salesmen. For these salesmen intelligence scores and production records were available [12]. For purposes of analysis, samples of 50 judges of each sex were taken, and estimated intelligence was correlated with measured intelligence for each judge. Then the correlation coefficients obtained with the 50 judges of each sex were averaged. The result was $-.22$ for men and $-.18$ for women judges. The situation is somewhat improved if the 50 judgments regarding a given salesman are pooled to get an average rank. These pooled estimates correlate with actual intelligence to the extent of $.62$ for men and $.33$ for women. The correlation of $.62$ might be of some significance in a personnel situation except for the inconvenience of finding 50 judges to examine the photograph of each individual applicant.

General Vocational Aptitude. Mention should be made of one investigation in which prediction of general vocational aptitude rather than a specific trait was attempted from photographs [13, 20]. The photographs involved graduates of an eastern institution who had majored in medicine, law, education, and engineering. Five of the most successful and 5 of the least successful in each group were selected with the cooperation of the alumni office. Photographs of these persons were available taken at approximately the time the experiment was conducted—namely, 25 years after graduation—together with photographs at the time of graduation. These photographs were submitted to groups of judges who were merely required to estimate whether the individual was a success or a failure. In one series the judges were college students; in the other they were employment managers and interviewers experienced in personnel work. Twenty-four of the former and 20 of the latter participated. The results are summarized in Table 6, which gives the percentage of correct judgments under the conditions indicated by the row and column. For example, with personnel people judging the photographs of graduates of 25 years' standing as to success or

TABLE 6. PERCENTAGE OF CORRECT JUDGMENTS OF VOCATIONAL SUCCESS FROM PHOTOGRAPHS⁷

	Older Photos	Younger Photos	Both
Personnel workers.....	53	52	52
Students.....	51	47	49

failure, 53 per cent of the judgments were correct; but 50 per cent would be expected by chance. Similarly with the other combinations the results indicate no tendency to be able to judge general vocational success from the photographs.

Special Vocational Aptitude. A few studies may be mentioned in which efforts were made to estimate specific aptitude from photographs. The investigation mentioned previously regarding intelligence of salesmen embodied the additional feature that the judges of the photographs were asked to estimate selling ability. It was possible to compare these estimates with the actual selling ability as indicated by production on the job. The correlations actually were slightly negative. Averaging the correlations obtained from 50 judges of each sex, we have $-.16$ for male and $-.17$ for female judges. When the individual ratings are combined into average ranks before correlating, the result is even worse, namely, $-.38$ for men and $-.22$ for women.

A somewhat similar technique was applied in the case of teachers [8]. Photographs of teachers were sent to superintendents, secretaries of school boards, and secretaries of teacher placement bureaus—24 pictures in all. These judges were asked to rank them in order of success. While there were some cases of agreement, nevertheless a given teacher always received all 24 ranks. These teachers were then ranked by six faculty members who went over the records on file in the placement bureau at the institution where they received their training. When these ratings by the faculty on the basis of all available data were compared with the ratings made by school officials on the basis of the photographs, there was complete disagreement. The correlations for

⁷ After Viteles, and Landis and Phelps.

different groups of judges were $-.10$, $.14$, $-.01$, and $.37$. In some instances the judges agreed fairly well with each other, but their combined judgments were widely divergent from the actual status of the teachers.

The Halo Effect. One other point that has wide implications in the whole theory of rating procedure may be noted in these experiments on judging miscellaneous traits from photographs. It is brought out by correlating estimates of various traits with one another to determine, for instance, whether persons who are rated high in humor are likewise rated high in perseverance, kindness, etc. Photographs of 20 women were ranked by judges with reference to six traits, and the average rank of each individual was obtained in each trait [7, 46]. By the use of these average ranks, each trait was then correlated with each of the others. The results are shown in Table 7. Any figure in the table indicates

TABLE 7. CORRELATIONS BETWEEN ESTIMATES OF DIFFERENT TRAITS MADE ON THE BASIS OF PHOTOGRAPHS⁸

Traits	Intelligence	Humor	Perseverance	Kindliness	Conceit	Courage
Humor47					
Perseverance88	.33				
Kindliness76	.65	.39			
Conceit28	-.03	.08	-.56		
Courage89	.43	.79	.72	-.25	
Deceitfulness . .	-.11	-.28	-.03	-.69	.66	-.49

the correlation between the trait listed at the left of that row and the trait listed at the top of that column. For instance, the correlation of humor and intelligence is $.47$, that of perseverance and humor, $.33$. It will be seen that humor, perseverance, kindness, courage, and intelligence all seem rather closely related. A person who looks as if he possessed a high degree of one of these appears as if he possessed a high degree of the others. Conceit and deceitfulness, on the other hand, show negative or low correlations with the above traits, but correlate highly with each other. These

⁸ From H. L. Hollingworth, *Vocational Psychology*, by permission of D. Appleton-Century Company, Inc., New York.

results suggest a danger in ratings of this sort—a factor that plays a role in ratings in general. There is a tendency for the judge to form a general impression that is favorable or otherwise, and to rate the person accordingly in a number of traits. This effect has been called a *halo effect*. This halo of general impression often colors estimates of various traits so that not much validity can be attached to the estimate of any one trait vs. another. The judge thinks he is evaluating the traits independently, but he is merely recording repeatedly his general impression. (Cf. also the discussion of the halo effect on p. 389.)

Another instance of the halo effect was brought out in the above-mentioned study of intelligence and selling ability as judged from photographs. There was a high correlation between estimated intelligence and estimated selling ability. The coefficients were .82 for men and .93 for women judges. Actually, however, the correlation between these two variables as measured by intelligence tests and sales records was $-.48$. Obviously the raters had gained some type of general impression and made their estimates accordingly.

The results of such studies as the foregoing are not encouraging to those who hope to predict personality from physiognomy. When estimates of mental characteristics made from photographs are compared with more certain criteria of those characteristics, such as the judgment of intimate acquaintances or measurements of intelligence, there are marked discrepancies between the two. An individual judge's results have little validity, and even when a considerable number of judges pool their estimates the results are far from what is to be desired. The only conditions under which it would be at all advisable to install such methods for employment would be where a corps of probably twenty or more persons were available to make these physiognomic judgments and average their findings. It is a question whether this procedure would be expedient. Inasmuch as scientific methods are available that do not necessitate the use of such a corps, it would seem wiser to devote one's effort to the use of these methods. They will be described in later chapters.

Evaluation of Commercial Systems of Character Analysis. Psychologists have been so busy improving their methods of using mental tests and other measurements for practical purposes of

employment that they have devoted little effort to experimental refutation of specific relations between aspects of physiognomy and mental characteristics that are assumed by commercial systems of character analysis. A few investigations of this sort, however, have been made and the results are presumably typical of what will be found if further alleged relations are studied.

Alleged Blonde and Brunette Traits. One of the most widely known systems of character analysis makes much of the mental differences between blondes and brunettes. As this is an easily observable anatomical distinction, it would be very convenient if character could be inferred therefrom. According to the system in question, this is possible, and a list is provided of the traits possessed primarily by blondes; a similar list is furnished for the brunettes. It was possible statistically to determine the validity of these lists [15]. Twelve "blonde traits," such as positive, dynamic, driving, aggressive, domineering, and 14 "brunette traits," such as negative, static, conservative, were arranged in a random order on a printed blank. These blanks were given to 94 persons who were above average intelligence. Each individual selected two pronounced blondes and two pronounced brunettes with whom he was well acquainted. For each of these acquaintances he went through the printed list of 26 traits and marked them with a plus or minus sign according to whether, in his judgment, the person possessed that trait or not. The people marking the blanks were not familiar with the particular system of character analysis involved and the traits occurred in a random order so that the alleged blonde ones would not be found grouped together. It was then possible to tabulate the percentage of blondes who were rated plus on the blonde traits and also who were rated plus on the brunette traits. These results are shown in Table 8. For instance, 81 per cent of the blondes are positive, which is an alleged blonde trait, but 84 per cent of the brunettes are likewise positive; 63 per cent of the blondes are dynamic, but so are 64 per cent of the brunettes. While brunettes are supposed to be negative, 17 per cent are found to be so in the actual results, but 16 per cent of the blondes are negative. The averages indicate that the 12 alleged blonde traits are possessed in general by 63 per cent of the blondes, but are also possessed by 61 per cent of the brunettes; the 14 alleged brunette traits are possessed on the

TABLE 8. PERCENTAGE OF BLONDES AND BRUNETTES RATED AS POSSESSING ALLEGED BLONDE OR BRUNETTE TRAITS⁹

Blonde Traits	187 Blondes	187 Brunettes
Positive.....	81	84
Dynamic.....	63	64
Driving.....	49	50
Aggressive.....	62	56
Domineering.....	36	36
Impatient.....	56	51
Active.....	88	82
Quick.....	70	68
Hopeful.....	85	85
Speculative.....	53	51
Changeable.....	53	43
Variety-loving.....	66	62
Average.....	63	61
Brunette Traits		
Negative.....	16	17
Static.....	28	31
Conservative.....	51	61
Imitative.....	39	40
Submissive.....	25	26
Cautious.....	54	60
Painstaking.....	56	61
Patient.....	43	52
Plodding.....	27	31
Slow.....	20	24
Deliberate.....	47	57
Serious.....	58	72
Thoughtful.....	67	70
Specializing.....	52	45
Average.....	42	46

average by 46 per cent of the brunettes, but also by 42 per cent of the blondes.

A somewhat different approach was made to this same problem by sending to 50 well-known sales executives a list of traits used

⁹ After Paterson and Ludgate.

in the system of character analysis referred to above [9, 244]. Each executive selected four highly successful salesmen and checked on this list of traits the ones they possessed. Results were available for 152 salesmen. The outstanding characteristics mentioned were: positive, dynamic, driving, aggressive, active, quick, painstaking, hopeful, patient, serious, thoughtful, specializing. Of these, seven are "blonde" and five are "brunette" traits. Obviously it would be difficult to select a good salesman on the basis of his complexion.

One other bit of evidence bears on this same question. The system alleges that persons of mechanical bent are typically of light complexion. In a survey of 400 metal workers, most of whom were presumably somewhat mechanically inclined, 16 per cent were light, 32 *per cent* dark, and 52 per cent medium [10]. Obviously there is no tendency for them to be typically light. The majority are medium, and there are more dark than light complexions in the group.

Miscellaneous Physiognomic Factors. From current systems of character analysis a considerable number of miscellaneous physiognomic characteristics were selected which were claimed to be an index of mental traits, and these physiognomic characteristics were actually measured [2]. The traits studied were the following: judgment, intelligence, frankness, ability to make friends, will power, leadership, originality, and impulsiveness. These traits were selected because there was fair agreement among the physiognomists regarding them. The physical measurements were made with calipers, sliding compass, steel tape, and head-square. The character analysts use only their eyes. These experimenters used instruments which must have made their measurements at the worst far more accurate than the character analysts' at their best. Persons intimately acquainted with the individuals who were measured provided estimates as to these particular mental traits. Furthermore, the individuals were placed on the stage before a group of judges who were unacquainted with them and they were estimated casually for the mental traits to determine the possibility of a practitioner being able by "intuition" to estimate traits in an interview, although he actually professed some physiognomic basis for his judgments. The experimenters measured a large number of physical charac-

teristics which the analysts claimed correlated with the mental traits above enumerated. Anywhere from 20 to 36 different items were measured in connection with each of the eight mental traits, making a total of 201 different measurements obtained upon each individual.

Thirty students were measured in this fashion and were rated as to the mental traits by members of their fraternities or sororities. These ratings prove to be quite reliable, i.e., the different members of the fraternity or sorority agree rather closely with one another in rating a given individual. These opinions of acquaintances thus form a pretty good standard by which to evaluate the physiognomic measurements. On the other hand, the reliability of the physiognomic measurements is low. For instance, a number of them are alleged to indicate judgment. If the relative standing of the students in one such set of measurements is obtained and correlated with their standing in another measurement which is supposed to indicate the same mental trait, the correlations are uniformly small. In other words, the theories of the character analysts with reference to physiognomic indications of a given mental trait are discordant among themselves.

The crucial point is, of course, the correspondence between the physiognomic measurements and the estimates made by close associates. The best way to summarize the entire results is to average all the correlations of the physiognomic measurements for a given trait with the associates' judgments of that trait. For instance, with intelligence 29 different factors were measured. Each of these is correlated with estimated intelligence. The average of these 29 correlations is then computed and, as indicated in Table 9, gives .03. Similar averages for the other mental traits appear in the first column of the table. It is obvious that these correlations are all extremely small and show practically no relation between the alleged physiognomic indicators of mental traits and the actual possession of those traits. The average correlations between the opinion of the casual observers and the physiognomic measurements are given in the next column and are likewise of insignificant magnitude. The results for the close associates and for the casual observers are correlated in the last column. These correlation coefficients are slightly higher than the others and might indicate a very slight possibility that the judges,

TABLE 9. CORRELATION BETWEEN RATINGS OF CLOSE ASSOCIATES AND PHYSIOGNOMIC FACTORS¹⁰

Traits	Close Associates and Physiognomic Measures	Casual Observers and Physiognomic Measures	Close Associates and Casual Observers
Judgment	-.01	.14	.32
Intelligence03	.05	.02
Frankness05	.15	.21
Friendliness	-.11	.19	.18
Will power	-.07	.04	.26
Leadership	-.04	.07	.31
Originality09	.08	.32
Impulsiveness10	-.07	.20

through "intuition" or something of the sort, are able to evaluate certain aspects of personality. However, only three of the traits yield correlations as large as .30; the other five are distinctly less. The general conclusion of the study is that "the average of 201 correlations between various physical traits purported to reveal variations in character traits and our criterion is .00 with the correlation varying from .00 as chance would account for. Physical measurements which underlie character analysis agree neither with themselves nor with other measures of character."

A somewhat related study developed a technique for making accurate head measurements on photographs which were taken by putting the subject in a standard apparatus, with certain key points marked at the time the photograph was taken [18]. Numerous measurements were made on each of 100 subjects, including such factors as length and breadth of head, length of face, length of nose, distance between pupils, and numerous ratios between various measurements. Altogether there were 27 different variables. They were correlated with estimates made by friends on five traits—sociability, perseverance, leadership, aggressiveness, emotional excitability. The correlations were all small; only one was as large as .30.

One experiment is available in which the practitioner of a

¹⁰ After Cleeton and Knight.

pseudo-scientific procedure applied his own methods to a group of individuals and made predictions which could be checked by objective criteria [5]. In this case a practitioner of "vitosophy" applied his technique to 20 university students and made estimates as to their ability in mathematics, written speech, science, general scholarship, general intelligence, mechanical ability, and musical ability. The students wore uniform laboratory coats during the examination. The vitosopher made various skull measurements by means of a rotating hemisphere centered at the ears, had the subject blow his breath against the hand of the examiner, inspected the back of the subject's hand, and examined the teeth. The examiner estimated the students' proficiency as grades A, B, C, D, or E in the variables indicated. As a criterion for the first four, grades received in the university classes were available. Intelligence was measured by one of the standard tests (Army Alpha). Mechanical and musical ability were estimated by the subjects themselves. The correlations between the five objective measures and the estimates ranged from .32 for speech to $-.21$ for intelligence, and averaged about .04. The correlations for the other two variables were $-.55$ and $-.31$. When all the variables were used, the average correlation was $-.08$. Obviously vitosophy was not very successful in its predictions. The practitioner, however, must be given credit for being willing to submit his technique to an experimental test. Most pseudo-psychologists who make an appointment for a similar test become ill on the appointed day.

Present Extent of the Use of Physiognomic Methods. It is difficult to ascertain to what extent methods like the foregoing are being seriously used for employment purposes. There is no doubt that many persons are using some popular or personal generalizations of this sort as a supplement perhaps to other criteria. A questionnaire was circulated in 1922 among 100 employment managers and insurance agency managers asking if they used any system of character analysis, and if so what one [11]. Sixty-five replies were received—22 from insurance men and 43 from industrial concerns. Two of the former and four of the latter stated that they used some system of character analysis. It is probably safe to say that six out of 100, rather than six out of 65, were using some system, because those who used one would

be more apt to reply than those who did not. Six per cent is not a large figure, but it is 6 per cent too many to be using this sort of method.

A more recent survey was made of 200 concerns in Connecticut and western Massachusetts in 1936 [14, 60]. They were asked as to whether they had tried any system of character analysis in connection with employment work. Of the 200, only two admitted having tried one or more systems; none admitted that they were using such a system at the time of the inquiry, and none of them expressed any confidence in any of these techniques. Perhaps the difference in the results of the two studies represents a whole-some trend.

In the light of the experiments on character analysis by means of physiognomy, one wonders why the methods should at present be in use at all—why the system has not already killed itself. The answer lies in the fact that some practitioners are able occasionally to hit the mark and make successful predictions or give valuable advice—ostensibly by the use of their system, but in reality on some other basis. In the first place, the analyst may hit by chance *one of the many* occupations for which the individual is fitted. It is not always a case of there being one job and only one in the world in which an individual may be successful; there are usually many lines in which he may achieve success. Consequently, selecting one of these by accident is not such a remote possibility. We often do this very thing ourselves; otherwise most of us would be maladjusted. The analyst can frequently by casual observation eliminate some possible lines of work for which the person obviously is disqualified and thus stand a greater chance of accidental success in predicting from the remainder. In the second place, the analyst may in the course of conversation discover likes and dislikes which may be of some vocational significance. He will perhaps be enabled with this information to make common-sense suggestions quite apart from any system. In the third place, if a person pays for vocational advice and the “expert” recommends a certain line of work, the individual will perhaps try harder than he would otherwise and hence reach a higher level of success than, with his ability, he would ordinarily attain. The “expert,” of course, gets the credit for this. Finally, when people are discussing such cases and com-

paring notes they are apt to slip into a common human fallacy of stressing the cases of coincidence and forgetting the others. This tendency to neglect the negative instances plays into the hands of the pseudo-psychologists. Persons remember the one case in which they hit the mark and forget the other ninety-nine times in which they miss. Scientific employment psychology may not always hit the mark, but it does so far more frequently than does pseudo-psychology.

SUMMARY

Before proceeding to the discussion of psychological methods in employment, it is necessary to clear the ground of a considerable amount of pseudo-psychology which is being widely commercialized and is masquerading under the name of psychology to the detriment of the real science. A number of these pseudo-psychologies have played a role in employment problems in recent years. *Astrology* has no scientific basis and its generalizations have not been evaluated statistically, but it is actually in use. *Spiritualism* has certainly nothing to contribute until its actual existence can be proved. It is illogical to assume communication with spirits until telepathy can be demonstrated, and this in the opinion of the great majority of psychologists has not as yet been accomplished under laboratory conditions. Yet spiritualistic mediums are consulted on various problems of a vocational nature. *Phrenology* started with the scientific findings regarding the functions of certain regions of the brain, but went far beyond the experimental results. It erroneously assumed a much more detailed localization of functions and a direct relation between the functional capacity of a brain region and its size, and it used a few casual observations as a basis for generalization.

Physiognomy is the most prevalent of these pseudo-psychologies. Our popular beliefs in it are due largely to the fact that one thing we see is associated with similar things (e.g., a short neck suggesting a bull and hence aggressiveness); to the fact that our observations are influenced considerably by what we expect to see (e.g., a weak handshake causing us to watch for further indications of vacillation); and to our assumption that, inasmuch as habitual activities often leave bodily traces (e.g., the round shoulders of the studious), it is logical to argue back-

ward from those traces to the activity in question. These popular beliefs, however, pave the way for our acceptance of commercial systems of character analysis from physiognomy. The validity of such beliefs and systems has been to some extent studied scientifically. Estimates of various mental traits made on the basis of photographs by judges who never saw the individuals themselves were compared with careful estimates of those same traits by intimate acquaintances or with actual measurements of the traits. The results indicate that a single judge is very inaccurate in making such estimates from physiognomy, and that while matters may be improved somewhat by using a considerable number of judges and averaging their results, the correspondence of this pooled estimate and actual possession of the trait is not sufficiently close to make the physiognomic factor of much practical value. The results are similar when judgments of general success in life or achievement in a specific vocation are made on the basis of photographs. Moreover, the results are often vitiated by the halo effect, or a tendency to get a general impression of good or bad and rate the person high in most desirable traits, or vice versa, instead of evaluating the traits independently.

These studies show the futility of the judgment of character traits from physiognomy when the judge is left to his own devices. The futility has been shown to be equally great when there is a question of the relation between specific physiognomic and mental characteristics claimed by commercial systems of analysis. The alleged relation of complexion to specific character traits is without foundation, for it has been shown statistically that blondes possess the traits that are supposed to characterize brunettes to just as great an extent as do the brunettes themselves, while the brunettes rival the blondes in the possession of the alleged "blonde traits." A group of intimately acquainted persons rated one another in several traits which have received considerable attention from the character analysts. Alleged physiognomic correlates of these traits were actually measured with calipers and steel tape—200 physical measurements upon each individual. These measures were separately correlated with the criterion provided by estimates of intimate acquaintances. The correspondence between the physiognomic measures and the actual traits is exactly what would have been expected by chance.

The physical measurements give no indication whatever of the mental traits in question.

The practical employment man is certain sooner or later to come in contact with some of these pseudo-psychologies, especially with the commercial systems of character analysis. From the foregoing considerations it is obviously to his interest to confine his efforts to scientific employment psychology rather than to invest in any of these psychology gold bricks.

REFERENCES

1. Anderson, L. D. Estimating Intelligence by Means of Printed Photographs. *Journal of Applied Psychology*, 1921, 5, 152-155.
2. Cleeton, G. U., and Knight, F. B. Validity of Character Judgments Based on External Criteria. *Journal of Applied Psychology*, 1924, 8, 215-231.
3. Coover, J. E. *Experiments in Psychical Research*. Stanford University Publications, 1917, 641 pp.
4. Eisenberg, T., and Reichline, T. B. Judgments of Dominant Feeling from Motion Pictures of Gait. *Journal of Social Psychology*, 1939, 10, 345-357.
5. Ford, A. A Check on Character Analysis. *Personnel Journal*, 1930, 9, 121-123.
6. Hollingworth, H. L. *Judging Human Character*. New York, Appleton-Century, 1923, 268 pp.
7. Hollingworth, H. L. *Vocational Psychology*. New York, Appleton-Century, 1916, 308 pp.
8. Johns, W. B., and Worcester, D. A. The Value of the Photograph in the Selection of Teachers. *Journal of Applied Psychology*, 1930, 14, 54-62.
9. Kenagy, H. G., and Yoakum, C. E. *The Selection and Training of Salesmen*. New York, McGraw-Hill, 1925, 244 pp.
10. Kitson, H. D., and Donham, C. L. The Personality of Workers in the Metal Trade—a Statistical Study. *Journal of Personnel Research*, 1924, 2, 460-466.
11. Kornhauser, A. W., and Jackson, A. W. A Note on the Extent to Which Systems of Character Analysis Are Used in the Business World. *Journal of Applied Psychology*, 1922, 6, 302.
12. Kornhauser, A. W., and McMurry, R. N. Ratings from Photographs. *Personnel Journal*, 1938, 17, 21-24.
13. Landis, C., and Phelps, L. W. The Prediction from Photographs of Success and Vocational Aptitude. *Journal of Experimental Psychology*, 1928, 11, 313-324.

14. Moore, H. *Psychology for Business and Industry*. New York, McGraw-Hill, 1939, 527 pp.
15. Paterson, D. G., and Ludgate, K. Blonde and Brunette Traits; a Quantitative Study. *Journal of Personnel Research*, 1922, 1, 122-127.
16. Rhine, J. B. *Extra-Sensory Perception*. Boston, Bruce Humphries, 1934, 169 pp.
17. Rogosin, H. M. Evaluation of Extra-Sensory Perception. *Journal of General Psychology*, 1939, 21, 203-217.
18. Sheldon, W. H. Ability and Facial Measurements. *Personnel Journal*, 1927, 6, 102-112.
19. Troland, L. T. *A Technique for an Experimental Study of Telepathy and Other Alleged Clairvoyant Processes*. 1917, 26 pp. (Published privately.)
20. Viteles, M. S. The Prediction of Vocational Aptitude and Success from Photographs. *Journal of Experimental Psychology*, 1932, 15, 615-629.

Chapter III

HISTORY OF SCIENTIFIC VOCATIONAL PSYCHOLOGY

The preceding chapter called attention to some of the pitfalls of pseudo-psychology which beset the practical man to the detriment of himself and of his attitude toward the real science. The perspective in which psychology as related to personnel is viewed may be still further enlarged by a consideration of its historical background.

INDIVIDUAL DIFFERENCES

Early Interest in General Laws. The early studies in psychology were directed toward the determination of general laws, whereas the differences between individuals are usually much more significant for practical purposes. Aristotle developed laws of association to explain why one idea calls up another; Weber and Fechner worked out the psychophysical law to express the relation between the intensity of the stimulus, such as light or sound, and the intensity of the sensation; Ebbinghaus derived certain laws pertaining to memory. While this type of work was of immense importance in laying the foundations of the theoretical science, it had little to do with sorting the applicants for a job. It was only after the groundwork had been partly laid and some psychologists turned their efforts from the general principles to the individual differences that progress was made in the field which is our present concern. A number of factors contributed to this shift of interest.

More Detailed Study of "Faculties." The earlier psychology made a good deal of the notion of "faculties" into which mind could be divided, such as the faculty of memory or the faculty of attention. It became evident, however, that these faculties must

be still further subdivided. It developed that memory for numbers and memory for words were two quite different things and that attention to a thunderclap and attention to an uninteresting book differed. In order to investigate these more minute differences it was necessary to arrange appropriate experimental material. Lists of words and lists of numbers were devised for the study of those two types of memory. Interesting and uninteresting materials were selected for the study of the different kinds of attention. This sort of material was the prototype of the mental test. Then, when experiments were conducted to determine the difference between memory for words and memory for numbers, it became obvious that, while such differences existed, there were also differences between individuals in their ability to retain the material. In the experiments on attention, while the expected difference in attention to interesting and to uninteresting material was obtained, there also proved to be differences between persons in the amount to which they could adequately attend. In this way the early experimental psychologists, attempting to subdivide such "faculties" as memory and attention, noted and became interested in these individual differences.

Need for Mental Measurements in Problems of Heredity and Education. On the other hand, practical considerations came from without the science to meet halfway the interest that was developing within. Galton and others were much interested in heredity. It was observed that many students who took honors at Oxford had parents who had done likewise; that one family had many lawyers and judges, while another had musicians and artists, among its ancestors; that some individuals with phenomenal memory had parents who excelled in that same respect. A good deal of data of this sort was collected, using qualitative estimates of the traits or abilities in question. Students of these problems came to realize the need for quantitative data and for some method of actually measuring the traits. Hence they looked to the psychologists for assistance in devising such measurements. Education was another field which early realized the need for psychology. It was observed that one child made rapid progress in school, while another was retarded. One individual sixteen years of age might be entering college, while another of the same age was still in the fourth grade. What caused this difference in

educational performance was a moot question. It led some of the pioneers to seek methods for measuring general ability or whatever mental factor was involved in school retardation.

Thus the interest of theoretical psychology itself in devising finer measurements for the study of general abilities, and the interest of those in other fields, such as heredity and education, in obtaining such measurements to use in their own problems, led to a considerable shift in emphasis from the general laws to the individual differences.

EARLY DEVELOPMENT OF MENTAL TESTS

Freshman Mental Tests at Columbia. The outstanding pioneer effort in the use of mental tests was at Columbia University in 1894. Under the direction of Cattell there was instituted a plan for testing the students in their first and fourth years. The purpose of the project is well expressed in the following paragraph which actually constituted the material for one of the memory tests:

Tests such as we are now making are of value both for the advancement of science and for the information of the student who is tested. It is of importance for science to learn how people differ and on what factors these differences depend. If we can disentangle the complex influences of heredity and environment we may be able to apply our knowledge to guide human development. Then it is well for each of us to know in what way he differs from others. We may thus in some cases correct defects and develop aptitudes which we might otherwise neglect [2].

The tests used were for the most part those of sensory capacity, such as color blindness, auditory acuity, perception of pitch, sensitivity to pain, or else measurements of the speed and accuracy with which certain tasks could be accomplished, such as marking 100 letters or making 100 movements. This project is typical of the early test work. Miscellaneous tests were devised and tried in order to see whether they differentiated persons from one another and to determine how a given individual scored with reference to the rest of the group.

The Standardization of Tests. Cooperation in the standardization of tests was the next step historically. After various workers had devoted considerable independent effort to devising miscel-

laneous tests and trying them out on small groups of individuals who were available, it became obvious that cooperative effort would facilitate matters. Some of this took place through the usual channel of publication in scientific journals of descriptions of material and methods for tests so that they could be tried by other investigators and results could be compared. In 1906 the American Psychological Association appointed a permanent committee to act as a general control committee on the subject of measurements. It was charged, among other matters, with the development of a series of group and individual tests. The committee functioned for several years and issued reports upon tests for auditory acuity, pitch discrimination, imagery, and a set of "association tests," involving such things as naming colors, canceling numbers, learning a code, giving opposites, and following complicated and confusing directions. These association tests were widely used for a time in their original form and served as a pattern for investigators who devised other tests along similar lines [8].

Binet. Another significant contribution to the development of tests was the work of Binet. His problem was to devise means of measuring intelligence of children. The method consisted essentially of finding a set of questions for children of each age such that the average child could answer them satisfactorily. Consequently, if a child was backward he would fail on the questions for his own age, although he might succeed in answering questions designed for some lower age. Binet began his work about 1900 and published his original intelligence scale in 1908 [1]. It was subsequently translated and revised by Goddard, Terman, and others, and is now, in revised form, one of the most widely used intelligence tests.

Whipple. In 1910 Whipple published the first edition of his *Manual of Mental and Physical Tests* [7]. This presented most of the important tests that had been devised and used to any great extent up to that time. They were classified, standards given as far as available, and considerable data presented on the relation of the tests to each other and to various factors such as age. This compilation of material and procedure for giving a considerable number of tests was valuable, as it enabled many per-

sons to give similar tests under similar conditions and to compare results.

This perhaps marked the high spot in the early development of tests for their own sake. The emphasis up to that point was largely upon devising tests and standardizing them on various groups of individuals. It was eminently desirable that the technique should go through these stages before being put into the practical situation. Efforts to use tests for hiring employees in 1894 would have been premature. Much had to be learned about the principles to be observed in the construction of test material, in the wording of directions, in the selection of time limits, and in the scoring of results. In short, the whole theoretical technique had to be reasonably well developed before it was profitable to apply the tests to practical ends.

COMPARISON OF TEST SCORES WITH OCCUPATIONAL ABILITY

The next step in the history of employment psychology was the comparison of efficiency in the tests with efficiency in the occupation. If those who were effective in the occupation made high test scores, and vice versa, the tests could then be used with applicants for a position to predict their future ability therein. The pioneer efforts in this direction were made by Münsterberg about 1911 with his study of motormen of the Boston Elevated Railroad [3]. His test consisted essentially of an endless belt arranged to pass under a small opening so that the person being tested might discriminate between different figures in different locations on this moving belt and react to them according to their significance. The novel feature was that the test was given to actual motormen and the test scores compared with their service records. It developed that motormen with a good record and few or no accidents made somewhat higher scores in the test than did those with a bad record of accidents.

Münsterberg also gave a series of tests to girls in a school for telephone operators. The tests themselves involved such abilities as memory for numbers, judgment of distances, rapidity of movements, and speed of association. The progress of the girls in the school was compared with their test scores and some tendency was manifest for those with satisfactory progress in learning the work of a telephone operator to make higher scores in the tests

than did those with unsatisfactory progress. The advance made in these studies was fundamental. Hitherto the tests had been standardized on anybody. Now they were standardized on persons engaged in a particular occupation, and efficiency in the tests was compared with efficiency in the occupation. This same procedure is basic to modern employment testing, namely, testing the tests. Statistical methods have improved, ingenuity in devising tests has increased, and many technical points have been perfected, but the general principle is the same.

Shortly after this time various other psychologists began to compare test scores with occupational criteria in similar fashion. Scott started his work on methods for selecting salesmen, comparing test scores with sales records [4]. At Carnegie Institute of Technology there was organized the Bureau of Salesmanship Research which embarked on a five-year program of cooperative research along these lines.

Then came the war.

PSYCHOLOGY IN THE FIRST WORLD WAR

Organization of Psychologists. On the day war was declared it happened that a group of psychologists from the eastern part of the country were gathered for one of the informal conferences, such as they often have, to talk over their problems. The discussion turned immediately to psychological war problems which it might be profitable to investigate and methods of approaching these problems. This was the first war conference of American psychologists. A few days later the officers of the American Psychological Association met and organized the psychologists of the country. The military problems were classified as far as possible and committees appointed. These were at first rather informal, but later operated under the National Research Council. They worked upon a wide variety of military psychological problems, such as the psychological examination of recruits, selection of aviators, gun-pointing, night-observing, training and discipline, incapacity, reeducation, emotional stability, self-control, propaganda, and tests for deception.

This is not the place to recount the work of all of these committees. Some were dealing with problems analogous to the employment problems of industry and some with entirely dif-

ferent problems. Only the former will be mentioned in the present connection. Suffice it that the war advanced applied psychology ten years in a few months. When it became absolutely necessary to do something, we found that there was much more psychology to apply than we had realized.

General Mental Examination of Recruits. One of the committees above mentioned took up the problem of the general mental examination or intelligence testing of recruits. It seemed plausible that different branches of the service and different ranks might require different degrees of general intellectual capacity. Accordingly, a group of psychologists who had previously been most closely in touch with problems of intelligence measurement undertook to devise a test which could be given to large numbers simultaneously, which could be scored by clerks who did not have psychological training, and which would yield a reliable indication of general mental ability. Prior to this time most testing had been individual, i.e., one person at a time was examined orally by a skilled examiner. It would obviously have been impossible for the available skilled examiners to examine individually a million men. Starting with the best available information regarding Binet and allied tests, a preliminary form was devised and tried out on a few thousand men. In the light of the results it was revised and developed into its final form, the Army Alpha test. It was ultimately given to 1,726,000 men. Its uses were many and varied, but some of them were similar to those encountered in current employment psychology.

For instance, there was the problem of eliminating from positions of responsibility those of such low mental status as to render them dangerous. In the Army some 8000 men were discharged because they were mentally unfit for duty; another 10,000 were assigned to labor battalions because of their low intelligence, and 9000 were sent to developmental battalions for observation.

Then there were problems of promotion, recommendation for officers' training camps, and the like. As a matter of fact, it was found that in the Army the average intelligence of the commissioned officers was higher than that of the non-commissioned officers, while these in turn excelled the enlisted men in average intelligence. This fact might be used in promoting men from the ranks. If the officers had higher intelligence, it seemed plausible

that a private of high intelligence, other things being equal, constituted better officer material than did a private of low intelligence. Army Alpha proved very useful in the post-war years because the items had all been carefully selected after experimentation and standards were available based on nearly 2,000,000 men. Many subsequent experimenters used this test in its original form. Others modified it by way of abbreviation or rearrangement of items. It served as the prototype for most of the post-war group tests of intelligence.

Selection of Aviators. Another committee took its cue from the fact that some recruits who passed all the medical examinations were a failure as aviators. Tests were devised for special capacities, such as speed of reaction, judgment of distance and velocity, ability to detect slight changes in equilibrium, and emotional stability. The test scores were correlated with ratings by flying instructors and with the number of hours of flying with an instructor necessary before the man was permitted to solo. It was possible to select a group of tests which would give a fairly good indication of aptitude for flying. The official installation of these methods was under way at the time of the armistice.

The special contribution of this work to the advance of employment psychology was the greater refinement of statistical methods. In addition to determining the relative importance of the different tests by correlation procedures, they were *weighted*. This involved ascertaining just how much importance should be attached to each separate test in the final combined score in order to get the best possible prediction of flying ability. It was found that certain tests overlapped, i.e., to some extent measured the same thing. For example, separate tests were devised for attention and speed of reaction, but the two were not necessarily independent. A man with good attention tended to be a little quicker in reacting because he paid closer attention while waiting for the signal. This became evident when the correlation coefficient between the two tests was computed. In similar fashion it was found that those who excelled in a memory test likewise were superior in an attention test, presumably because paying better attention facilitated learning the memory material. Here again the two tests overlapped and correlated appreciably with each other. It was then necessary to make allowance for this

overlapping of tests—otherwise one particular trait such as attention might receive undue importance or “weight” in the combined score. This allowance was made by partial correlation—finding the extent to which a given test correlated with flying ability when the effect of the overlapping tests was statistically eliminated. This technique will be discussed later (Chapter VIII). It had been developed previously, but this was one of its first applications in a practical vocational problem.

Soldier's Qualification Card. The committee on classification of personnel dealt with a group of specific employment problems. Something like half the men in the Army have to ply some special trade, and it was obviously advantageous to assign to such duty some man who already had ability in the trade involved. The problem then was to discover in the draft the men who had pursued these trades and make them available. One method of approach was to obtain information in the preliminary interview with the recruit, systematize it, and incorporate it in some standard form. Study of this problem led to the soldier's qualification card. The recruit was interviewed with reference to his personal history, occupational experience, education, etc. These items were entered on a standard card; standard terms and symbols were used. These cards were then tabbed at the top, the position of the tab indicating the trade with which the man was familiar and its color indicating his proficiency as far as it could be ascertained. When men of a certain type were needed, it was thus possible to look through the files of a unit and select in a few moments, by following down the tabs in a certain position, the men in that unit who were proficient in the trade in question.

Trade Tests. The above procedure had one drawback. In the interview a man would often make false claims as to his ability in a trade. As a matter of fact, it developed that about 30 per cent of those who claimed trade ability were totally inexperienced in the trade in question. If a man was assigned to duty involving carpentry on the basis of his own statement and could not drive a nail, the efforts of the interviewer, clerks, and others were wasted. Hence the trade test was developed actually to measure the man's trade ability—whether he was an expert carpenter or a journeyman or apprentice or whether he was a mere novice. Sometimes standard questions were asked about tools and

materials and processes involved in the trade—questions such as an experienced tradesman should be able to answer. These trade tests were evaluated by comparing scores with known ability in the trade. The standards were obtained in various industrial plants with men who had an actual trade status.

The trade test opened up another aspect of employment psychology. Hitherto most of the efforts had been devoted to predicting the aptitude or potential *capacity* of a workman to be successful in some particular job after due training. The trade test measured *proficiency* which the applicant possessed at the time of the test rather than any future possibilities. While the tests of capacity play the larger role in industry, the trade test has its place. The Army was the first situation where it was developed on any considerable scale.

Rating Scales. Army personnel work led to the consideration of certain mental traits, such as leadership, character, or general value to the service that could not be measured objectively. This problem was particularly important in dealing with officers. In the past a certain evaluation of such factors had been made, of course, in considering cases of promotion. But one officer in rating his subordinates would often use entirely different standards of judgment from those used by another officer and would attach different importance to different traits. The committee consequently found it desirable to develop a systematic rating scale covering certain specific qualities. It was ascertained as the result of appropriate study and evaluation of questionnaires that a limited number of qualities or traits were outstanding in the successful officer. These qualities were carefully defined and their relative importance was ascertained. A scale was then arranged with the maximum and minimum value to be assigned to any quality fixed. The procedure consisted essentially of selecting from well-known officers the names of a few individuals who possessed a given trait, such as leadership, in high, low, or average degree and assigning them standard values on a "master scale." The subordinates who were being rated were compared with this "master scale"; man-to-man comparisons were made and the subordinate was assigned the numerical value attached to the officer on the master scale most similar to him. This officers' rating scale was one of the first attempts at systematic develop-

ment of a technique for estimating scientifically these non-measurable mental characteristics. Thus psychological methods underwent a considerable development during the war, and inasmuch as many of the problems undertaken were of the type epitomized by "the right man in the right job," this work played an important part in the history of employment psychology.

POST-WAR PSYCHOLOGY

The beginning of 1919 found psychologists more interested than hitherto in personnel problems, and numerous individual research projects, particularly dealing with tests for employees, were launched. Considerable effort was devoted to the perfection of further tests and their validation in employment situations. Various occupational groups were studied as the occasion arose. In fact, some enthusiasts erred in overselling the field. This was especially true of persons with limited psychological training who had engaged in some military personnel work during the emergency and felt that they were well qualified. They attempted personnel projects which were not feasible or at least were beyond their capabilities, and when their results did not come up to expectations, psychology was blamed. It took some time to live down a number of unfortunate experiences of this sort. However, adequately trained psychologists with the background of their brief military experience went into various employment departments as members of the staff—one in a munitions plant, another in a rubber tire factory, another in a silk mill, another in a department store, and several in offices employing large staffs of clerical workers. Some projects organized at that time are still in progress.

As a number of psychologists made connections with industrial concerns and set up departments to do scientific personnel work using psychological techniques, it was natural that other individuals should attempt further projects along this line on a consulting basis. Usually these persons had some other connection, presumably academic, but occasionally they went into this work on a full-time basis and attempted to build up a practice. Advanced students in industrial psychology have frequently served internships in industrial concerns.

Cooperative Research. Much progress has been made through

the cooperation of various individuals or groups on research problems. One type of cooperation involves the interchange of results and methods between psychologists. It is a rather common practice, when one has completed a project dealing with employment or some other industrial problem, for him to publish his findings in a scientific periodical so that others may have the advantage of his experience and also so that others will not duplicate his experiments.

In a more specific type of cooperative research, a number of business concerns and scientists work together on a particular problem. For instance, the turnover among salesmen is of concern to business men and of interest to psychologists. The latter, however, may be occupied with their own work and unable to spare sufficient time personally to study the turnover problem. In such cases it has proved feasible for the business groups to contribute financially to the support of an organization that will undertake this research problem. The scientists who are interested can supervise the more detailed work carried on by a staff hired for the purpose.

Typical of this cooperative research is the work of the *Bureau of Salesmanship Research* that was organized at the Carnegie Institute of Technology. The head of a large insurance firm came to the Institute with a request for courses in salesmanship that went somewhat further than the conventional course. His attention was called to the need for more facts, such as the differences between successful and unsuccessful salesmen, their aptitudes and traits, various kinds of appeals, and methods of selecting men and providing incentives. As a result of this conference other firms were approached, and finally about thirty concerns contributed over a period of years to the support of the Bureau of Salesmanship Research that was thus established. A competent staff was organized and embarked on a systematic study of salesmanship. In addition to contributing funds, these companies made their records and their experience available to the research workers so that all such information was put into a common pool. They furthermore cooperated in carrying out experiments with various groups of salesmen and with different methods. The Bureau was governed by a board representing both the Bureau and the cooperating concerns. This is not the place to recount

the work of this Bureau; it is cited merely to illustrate this type of cooperative research. Although its work was interrupted to some extent by the war in 1918, the Bureau developed a series of "aids" for sales managers consisting of model application blanks, model letters of reference, various improvements in the interview procedure, and batteries of tests for selecting salesmen. These "aids" were distributed to the cooperating companies.

Other similar bureaus were the outgrowth of this one. For instance, one was organized to meet the problems of local retailers. It prepared employment tests, trained members in specific methods of correcting difficulties, and studied sales personalities. Results in retail stores were checked by "service shopping" in which certain individuals were hired to make purchases incognito and to take careful notes on what happened in each sale. This service shopping gave a quantitative expression of the percentage of dissatisfied customers, and statistics showed how this percentage decreased as the result of the bureau's work.

A similar cooperative project is represented by the *Life Insurance Sales Research Bureau*. This Bureau is financed by some of the insurance companies and the staff deals with such problems as more effective methods of selecting insurance salesmen. For example, a rather extensive study was made of the items on the personal history blank with a view to determining their validity and weighting them accordingly. The results of such studies are made available to the contributing concerns.

The *National Research Council* was organized under federal charter of the National Academy of Science and comprises various subdivisions, among them a Division of Anthropology and Psychology. The Council is not merely a laboratory or a repository of findings; it endeavors to coordinate research and further the organization and support of undertakings which demand the cooperation of individuals or institutions or both. Some of these projects have been in the personnel field. A typical one involved collaboration with the Civil Aeronautics Authority on the selection and training of pilots. A considerable number of projects were farmed out to various academic institutions, with a few scientists at the institution serving in an advisory capacity, the actual work being done by persons on fellowships or other stipends paid by the Civil Aeronautics Authority through the

National Research Council. This project contributed notably to scientific information about personnel, both civilian and military.

The *Personnel Research Federation* arose through cooperation between the National Research Council and the Engineering Foundation. Its membership comprises many agencies and institutions such as universities and business concerns, and many private individuals. Its purpose is to "further research activities pertaining to personnel in industry, commerce, education, and government wherever such researches are conducted in the spirit and with the methods of science." One of its important contributions is the publication of an official organ, *Personnel Journal*, through which many studies in this field are made public.

The *Psychological Corporation* was founded in 1921. It is incorporated not for profit, but for "the advancement of psychology and the promotion of the useful applications of psychology." It can pay no dividend greater than 6 per cent per year. All the stock is subscribed and held by psychologists, with the provision that at any time the American Psychological Association (one of the official national organizations of psychologists) can purchase all the shares, in this way bringing the Corporation under the Association's control. The original board of directors was rather unique in that every member was a psychologist of note and appeared in *Who's Who*.

One of the Corporation's initial objects was to serve as a contact between psychologists and the public. When a business man had a problem that was psychological in nature the Corporation stood ready to consider it and refer it to a reputable psychologist who was qualified to deal with it. The idea was to prevent these clients from coming into contact with a pseudo-scientist who would do them more harm than good. As the Corporation has developed, its activities have taken various other directions. In one of these, marketing research, it employs the interview technique, asking a sample of the public about such problems as their preference for certain brands of a commercial product or why they purchase a particular product. This information is useful in planning marketing policies. In personnel research, another activity, the Corporation develops tests or similar procedures for selecting employees along the lines discussed in this book. Other problems of management and industrial efficiency are sometimes

considered. Some individual service is rendered—clinical examining, counseling, and vocational guidance. Another function is performed by the test division which serves as a publishing, sales, advisory, and research agency for psychological tests and related materials. Considerable emphasis should be placed upon this advisory aspect. The Corporation refuses to sell tests unless it is convinced that they will be used scientifically. This organization has made a notable contribution in keeping applied psychology, particularly in its personnel aspects, on a scientific footing.

THE DEPRESSION YEARS

Several projects growing out of the depression have some bearing on the development of personnel psychology. One of these began at Minneapolis in 1931 when the Minnesota Employment Stabilization Research Institute was organized [6]. It was set up to study the broad question of unemployment in that community. As part of the program, studies were initiated with a view to determining the psychological characteristics of the unemployed and discovering what kind of people were dropped first in a depression. Other aspects of the program involved economic factors. As part of the personnel aspect, a large testing program was inaugurated to study the vocational aptitudes of unemployed persons. Many of those who came to the Institute to register were given rather extensive examinations looking toward their own guidance and subsequent supplementary training as well as contributing research data. Out of this Institute came a number of worth-while studies—for example, those dealing with clerical aptitude and with interest inventories. Some of the staff that worked at this Institute subsequently did similar work in the federal organization.

Another important venture is the United States Employment Service. Its major function at the outset was to register people who sought work and attempt to put them in contact with prospective employers. It has branches all over the country. In the period from 1933 on, from four to nine million persons were registered in the Service in any month. It became increasingly apparent, however, as individual cases were studied, that a great many of the registrants were young people without previous work experience. Consequently, the placement problem was not

the usual one of putting a person in a job related to one that he had already held. This threw the question squarely back to the consideration of the individual's capacity or aptitude for a given type of work—in other words, his potentialities. This is a problem that is central in much of the present psychological personnel work and it is natural that the Service should address itself to this particular problem. Thus was organized the Occupational Research Program [5]. A technical board was set up to guide the work and some of the Foundations contributed funds in the early stages of the program. An early emphasis was the development of adequate information about a large number of jobs, which ultimately resulted in the *Dictionary of Occupational Titles* comprising some 18,000 jobs. Other efforts went into the development of counseling techniques and the measurement of aptitudes or capacities for various kinds of work. Projects were established at various centers sometimes in cooperation with local communities or industries, and statistical work and analysis was coordinated through the Washington office. Studies have been made of various types of clerical and sales work and numerous trade tests have been perfected. The program has been especially fruitful in contributing readily usable forms and techniques for research in and administration of personnel work.

Other services similar to the above have been organized locally. Some of the larger cities sponsor organizations that serve the functions of registering applicants, interviewing, often testing, and classifying them. Local industries find these agencies useful in providing "pre-tested" applicants for positions.

MILITARY PROBLEMS AGAIN

In the emergency conditions beginning in 1940, especially after Pearl Harbor, personnel psychology received another impetus. On this occasion, however, there were few outstanding developments in the way of new techniques and procedures. Whereas in 1918 psychologists were breaking new ground, in 1941 they were mainly applying standard procedures in a new context and with new materials. Hence only passing mention will be made of the contribution of psychology in this period.

At the outset each national organization of psychologists designated one member to serve on an emergency committee that

would act as liaison between the scientists and the government. In the summer of 1940 a small committee developed the General Classification Test to serve somewhat the same function as the old Army Alpha. It is to some extent an intelligence test like the earlier one, but it also includes items in the field of mechanical aptitude. The trade tests devised by the U.S. Employment Service, now under the Federal Security Agency, were made available for military use. In aviation some work on training and rating pilots had already been done under the auspices of the Civilian Aeronautics Authority. One contribution, for instance, was the development of a standard flight in which a pilot was rated on a standard series of items. Some of these procedures were taken over into the military situation. Reserve officers with psychological training were called to active duty. Many psychologists worked under the Civil Service, others went through training and received commissions as military psychologists, and some of the more experienced were commissioned directly in the armed forces.

The long-time result of this project, aside from its military contribution, was not so much an advance in methodology as the training of a considerable number of psychologists in personnel procedures. As private industry has needed experts in this field they have become available. And many persons who became interested in it by virtue of their war experience have been active in promoting subsequent work along these lines.

Personnel psychology has reached the stage where its fundamental principles and techniques are fairly well established. Progress now consists of broadening the field and establishing norms on a wider range of occupations. An indication of desiderata for future progress is given in Chapter XVI.

SUMMARY

The early interest of psychology was in general laws. The shift to the consideration of individual differences came about through theoretical interest in analyzing various mental factors in more detail and through the need of those working in other fields, such as heredity and education, for a technique of mental measurement. The first extensive testing program was attempted

at Columbia in 1894. Subsequently, there was cooperation in developing and standardizing a variety of tests. A distinct contribution to the methods of measuring general intelligence was made by Binet and the rapidly growing body of tests for special capacities was collated by Whipple. The next step after the development of tests for their own sake consisted in comparing individual efficiency in tests with efficiency in an occupation. Münsterberg was the pioneer in this field with his experiments on motormen and telephone operators.

During the war in 1917-18 the psychologists experimented upon many problems of a vocational nature. The general mental examination of recruits resulted in a group test of intelligence that was the prototype for many subsequent scales. It also taught something about the occupational significance of general intelligence. The work of selecting potential aviators gave insight into the statistical possibilities of weighting a group of tests in order to predict vocational ability. The various qualification cards and blanks devised for Army personnel work have been useful patterns for subsequent personnel blanks. The trade test methods called attention to a new field—the measurement of proficiency as contrasted with capacity. The rating scale technique gave a method of obtaining quantitative data regarding traits that are not directly measurable.

In the post-war years the interest of psychologists in personnel problems continued. Some have been attached full time to industrial organizations and some have done individual consulting. Cooperative research projects were initiated and public or private agencies devoted some of their efforts to scientific personnel work. The depression was instrumental in initiating some research and service projects, notably the U.S. Employment Service. In the military emergency beginning in 1940 the psychologists again played a role, contributing the General Classification Test, personnel techniques in aviation, numerous trade tests, and men commissioned as military psychologists.

REFERENCES

1. Binet, A., and Simon, T. Le développement de l'intelligence chez les enfants. *L'Année Psychologique*, 1908, 14, 1-94.
2. Cattell, J. McK., and Farrand, L. Physical and Mental Measure-

- ment of Students of Columbia University. *Psychological Review*, 1896, 3, 618-648.
3. Münsterberg, H. *Psychology and Industrial Efficiency*. Boston, Houghton Mifflin, 1913, 316 pp.
 4. Scott, W. D. Scientific Selection of Salesmen. *Advertising and Selling*, 1915, 25, 5-6, 94-96.
 5. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
 6. Stevenson, R. A. The Minnesota Unemployment Research Project. *University of Minnesota, Employment Stabilization Research Institute*, 1931, 1, 1-26.
 7. Whipple, G. M. *Manual of Mental and Physical Tests*. 2 vols. Baltimore, Warwick & York, 1915, 688 pp.
 8. Woodworth, R. S., and Wells, F. L. Association Tests. *Psychological Monographs*, 1911, 13, Whole No. 57, 85 pp.

Chapter IV

TYPES OF MENTAL TESTS

Like any technician the personnel psychologist cannot operate satisfactorily without tools, and the mental test is his most frequently used instrument. In some instances, to be sure, he has to employ rating scales to secure in systematic form the judgment of persons familiar with the applicants. On other occasions he has to extract such scientific information as he can from a personnel or application blank. But, wherever possible, he resorts to tests because they are objective and quantitative. He is usually as loath to hazard a diagnosis of mental characteristics of a prospective employee without tests as a physician would be to diagnose bronchitis without the use of a stethoscope.

The subject¹ is sometimes tested orally, sometimes with blanks on which he writes or marks, sometimes with objects such as puzzles, and sometimes with simple mechanical contrivances which he manipulates. In all cases, however, the aim is to measure some capacity or proficiency in order to predict what the individual will do at some future time and under certain circumstances—for instance, when learning a particular job. It is not possible, of course, to measure the entire capacity in question any more than it is possible for a manufacturer to evaluate carefully every pound of wool in a shipment. In the latter case, however, the usual practice is to take some samples, examine them, and assume that the entire shipment is like the samples. A similar procedure is followed in mental testing. The ability to concentrate for a few minutes on a test blank is regarded as a reliable sample of the individual's ability to concentrate for a prolonged period on his daily task; the average speed with which one operates

¹ In psychological terminology the word *subject* denotes the person on whom the experiment is being performed or who is taking the test.

a telegraph key when a light flashes is taken as indicative of his quickness of reaction when driving through traffic; a sample of memory ability as manifest in a brief test is assumed to be typical of the person's memory for the details of his business. Another feature that characterizes most of the better tests is the quantitative nature of the results. The individual's score is expressed not as "good," "average," etc., but as a certain number of points. A mental test then may be roughly described as a scientific device for measuring quantitatively a typical sample of mental or motor performance in order to predict what an individual will do under certain circumstances.

Two things are rather essential in the psychologist's preparation for employment work. He must be familiar with the technique of test administration, that is, he must know how to use his tools just as a carpenter must know how to manipulate a saw. Mental test technique is the subject of the next chapter. But the psychologist also needs to know what tool to use on a particular occasion. It may be as ineffective for him to use a test of memory in order to predict ability at operating a hand-feed dial machine as for a carpenter to hack off the projecting end of a joist with a hammer. And just as we should consider a man a poor carpenter for attempting to smooth a plank with a chisel in ignorance of the fact that planes were available which would do a much better job, so a psychologist lays himself open to a similar charge of inexcusable ignorance if he uses an archaic and unreliable mental test when better ones are available. Numerous tests have been devised and perfected in recent years and descriptions of them are scattered through the scientific literature. Some of them are published by book companies or manufactured by companies that supply scientific instruments. Mention may be made of a bibliography by Hildreth [8] and a considerable collection of tests published in what is essentially a manual by Garrett and Schneck [4]. A person entering upon a project in employment research may find that some of these tests suit his purpose, or at least that they will afford valuable suggestions for developing his own tests.

An employment psychologist thus needs a familiarity with a considerable range of mental tests that have been developed in various connections. Some of these tests will be illustrated in

the present chapter. This discussion, however, will not constitute a miniature manual of tests. None of the examples will comprise a complete test; merely enough items will be presented to illustrate its nature. Neither will standards nor the relation of tests to occupations be given in the present connection. It is usually necessary, anyway, to recalibrate the tests in the particular employment situation that is under consideration. The effort will be merely to give the reader an idea of the types of mental tests that are available for the psychologist dealing with personnel problems.

CLASSIFICATION

Capacity or Aptitude vs. Proficiency or Achievement. Distinction has already been made between measures of capacity and measures of proficiency. The former, frequently called aptitude tests, measure primarily innate or hereditary factors, whereas proficiency or achievement tests, as they are sometimes called, are concerned with acquired abilities. It is not desirable to push this distinction between innate and acquired aspects to the bitter end. We speak, for example, of a memory test which we consider primarily a measure of native capacity. A person probably inherits something fundamental which gives him a greater facility in retaining materials. But this facility may be influenced by things that have happened to him. He may have developed better habits of attention which will help him to concentrate on material he is memorizing so that he will make a better score, not by virtue of superior retentiveness but rather because of a better attitude toward the material. It is difficult to disentangle inheritance from acquisition in situations of this sort, and practically it is not necessary to press the point to the limit. The distinction is fairly clear in the employment situation. The capacity or aptitude test is one that is designed to predict ultimate success in some kind of industrial performance in which the applicant at the time of testing has had no experience, while the proficiency or achievement test purports to measure his trade skill or the occupational ability that he possesses at the time of application. A common type of proficiency test, which, however, has a small role in industry, is the standard educational test that measures proficiency in school subjects such as history or geography. Trade tests will

be discussed in a later chapter. Some typical aptitude tests will be described in the present chapter as they are more ubiquitous and may be used for vocational prediction in many different situations. For purposes of discussion at least, we shall separate the capacity from the proficiency test. We shall see later that even with the former some account is taken of possible modification of the individual's test score by experience. There is a moot question as to whether practice significantly affects the score in a given test and a further question as to whether previous industrial experience or any life experience appreciably influences the score.

General vs. Special Capacity. Tests of innate capacity may be further classified into those involving special capacity and those requiring general capacity. There are situations in which a workman needs some rather special capacity, such as good attention or memory, quick reaction time, or accurate judgment of distances, in order to achieve success in his work. There are other situations in which no outstanding special capacity like this seems necessary; the person merely needs to be up to a certain general intellectual level, to be generally alert and able to adapt himself to circumstances—the thing that is often called intelligence.

TESTS OF SPECIAL CAPACITY

Practical Justification of Terminology. It is rather common practice in dealing with special tests to speak of them as tests of attention, tests of memory, and the like. This does not mean to imply, however, that the mind can be divided into clean-cut categories of this sort or that a test can be devised which samples one of these categories to the exclusion of all others. While such terminology may be undesirable for theoretical purposes, it is justifiable for the practical employment situation. The practical man has a better notion of what the investigator is driving at if he speaks of measuring the clerk's attention or speed of decision than if he discusses Test A and Test B. The personnel psychologist is not evolving a theory of attention or of judgment and it is not necessary for him even to define the terms which he uses. He is simply selecting certain tests that measure some aspect of mental performance, and the crucial point is whether these tests will enable him to predict an applicant's ultimate voca-

tional success. He can call the particular tests used anything he wishes without affecting their utility, but he usually gives them a name that has a definite connotation to most persons and that probably has *some* relation to the thing actually measured by the test.

The use of plausible terminology is helpful not merely in promoting a test with the management and making it seem reasonable to them, but also from the standpoint of test administration in securing adequate rapport with the subject being examined. Such things as terminology or apparently irrelevant external features of the test may contribute to the subject's attitude. For example, a British bakery concern used some tests of the form board type where blocks of different shapes had to be fitted into holes in a board that were of corresponding shape. The workers who were tested did not cooperate very well because they saw no connection between baking and fitting wooden blocks into holes. Someone hit upon the idea of painting the blocks and sanding them so that their texture resembled cookies. Thereafter rapport was much more favorable because the workers were now placing queerly shaped "cookies" in the holes in the board. Trivial things like this may make quite a difference in test administration. By the same token a plausible name for a test such as "We are now going to test your attention" will promote a more favorable attitude than "This is Test 17B." Some psychologists go so far as to eliminate the word "test" altogether and call it, for instance, a "work sample."

It is probably impossible to devise a test which measures one mental aspect to the exclusion of all others. Calling a test a memory test does not imply that it measures memory exclusively. If a person hears a list of words and then tries to reproduce the list, his efficiency will depend not only on his memory but on the extent to which he pays attention to the original reading. But this test will obviously involve memory to a greater extent than will a test in which the subject crosses out every letter A on a printed page. Thus, if a job rather patently necessitates good memory for its successful performance, it is desirable to try out a "memory test" which will probably measure that ability better than will a "decision test." Hence in the following discussion of tests for special capacity under different class headings, it must

be remembered that these headings are used merely for practical convenience, and that tests do not measure exclusively the thing indicated, but simply emphasize it more than other things. After all, the real problem is the correlation of test score with vocational ability regardless of what is actually measured by the test or what the test is called.

In the following pages a number of the categories of special capacity rather extensively used by employment psychologists are given, with one or more examples for each category. The list of categories is not intended as exhaustive and only enough examples are given to illustrate the variety of tests in use. Where specific time limits or the quantity of items constituting the test are mentioned, this is not intended as an arbitrary suggestion, but is stated merely for illustrative purposes. A person working in this field will usually go to original sources for his test material or else devise his own along lines suggested by the work of others.

No consistent effort will be made to indicate the originator of a particular test. In most cases it would be difficult because tests have been modified repeatedly since their origin and have appeared in scientific literature to such an extent that they are practically common property. A few tests that have been published under copyright are identified by the name of the author.

MOTOR CONTROL

Many industrial operations involve coordination between eye and hand. There are two aspects of motor control that are of significance, preventing motion—i.e., steadiness—and making motions accurately and rapidly. Lack of the former would seriously handicap, for instance, a jeweler assembling a small watch, while ineffectiveness in the latter on the part of a telephone operator inserting plugs in jacks would lead to more wrong numbers.

Example 1. The conventional steadiness test makes use of a metal plate pierced with round holes ranging in diameter from $\frac{1}{2}$ to $\frac{7}{64}$ of an inch. A needle or a piece of small wire is mounted on the end of a wooden rod so that the subject can hold it like a pencil and insert the point in the holes in the plate. The needle and plate are connected in series with a battery or transformer-rectifier and an electric counter so that when the needle touches

the plate the circuit is closed and the counter registers. The subject tries to hold the point of the needle in each hole for a prescribed number of seconds without touching the edge, beginning with the largest hole and working toward the smallest. The arbitrary number of the smallest hole he negotiates perfectly may constitute his score. Or an interrupter may be connected in the circuit so that whenever the needle is in contact with the plate an electric counter will record five times a second. The essential point is that a subject who is capable of preventing undue motion of his hand will make a better score whatever the method of administration.

Example 2. For indicating speed and accuracy of coordination a board is provided with three metal disks $\frac{1}{2}$ inch in diameter mounted at the corners of an equilateral triangle 4 inches on a side. The subject holds a stylus (a metal-pointed handle similar to a large pencil) with which he taps the disks in succession, going around the triangle repeatedly in one direction. Each tap records electrically on a counter. The examiner can easily note the number of circuits of the triangle and measure the time with a stop watch while the counter records the actual number of electrical contacts made. The subject may go as rapidly as he can and be scored according to his attempts and correct responses, or he may be required to keep time with a metronome while the number of errors is noted.

Example 3. A test of finger dexterity employs a considerable number of small metal pegs about the diameter of a match and a board with 100 or more holes in it. The test consists of placing the pegs in the holes. In some cases each hole is approximately the same size as a peg so that one peg goes in each hole. In other cases the holes are somewhat larger and three pegs have to be picked up simultaneously and placed in each hole. In another variation of this test the pegs are manipulated with a pair of tweezers rather than with the fingers.

Example 4. Some tests of manipulation with larger objects are typified by placing and turning tests. A board contains some 50 circular holes a little over an inch in diameter, arranged in 4 rows. A corresponding number of circular blocks are provided and arranged in a pattern on the table in front of the board. The test consists of picking them up as quickly as possible seriatim

and placing them in the holes. The blocks are deeper than the board and project above the surface so that they can be easily grasped. A supplementary test consists of removing each block, turning it over, and putting it back in the hole. The two sides of the block are painted a different color to facilitate checking the accuracy of the performance.

Example 5. Other coordination tests involve some type of pursuit. A metal plate may be substituted for the record on a phonograph chassis. A disk of insulating material an inch in diameter is set in this plate near the margin and flush with the surface. The subject holds a stylus similar to a small hammer with a hinge in the handle. The head is weighted so that gravity keeps it in contact with the metal plate. The purpose of the test is to hold the stylus in contact with the insulating disk so as to keep the circuit broken. Whenever the subject gets off the disk an electric counter records ten times a second. A more elaborate test of this sort involves a moving target driven by a complicated gear arrangement which goes through a very irregular pattern. The subject, with a stylus similar to that just described, "rides" this moving target. Whenever he gets off the contact the apparatus stops so that he can get on again. His task is to ride as far as possible in a given length of time.

Example 6. The tests just described involve primarily the muscles of the arm and hand. On occasion it may be desirable to test coordination of the larger musculature from the "feet up." In one such test a pendulum swings above a sink and from the end of it flows a small stream of water. The subject has a can with an aperture $\frac{1}{2}$ inch in diameter; standing in front of the pendulum he attempts to catch as much water as possible in a designated number of swings. Success in the test depends on using the legs as well as the arms and swinging the body back and forth to synchronize with the motion of the pendulum.

Example 7. A test involving more active use of the feet employs a mat of heavy rubber about a yard square with 25 small targets consisting of circles an inch in diameter painted on it. Underneath the mat is another unit such that stepping on any one of the targets closes an electric contact. In front of the subject at a convenient distance is a panel comprising 25 small lights in a square pattern. When two of these are illuminated the subject

must step on the two foot contacts corresponding. An electromagnetic ratchet stepping device controls the lights so that when the subject steps on the two correct contacts he immediately gets the next pattern of two other lights and must change his feet to new positions. The test can be performed by a stepping motion or by an actual hop with both feet simultaneously. The electric circuit is tapped so as to record on a counter the number of correct responses per unit time.

In some research studies effort has been made to determine an individual's natural tempo in motor performances like some of those described above. Subjects have been instructed, for example, to tap at their "most convenient" rate. While occasionally these rates have been found to correlate with some vocational criterion, they would not be practical for most industrial purposes because it would be difficult for an applicant for a job to operate at his natural rate. He would assume that a higher speed of performance was more apt to get him the job and hence he would not reveal his normal tendency.

Many tests of motor control like the foregoing require mechanical equipment of varying degrees of complexity. That is one unfortunate drawback of such tests. We shall note later the advantage of having tests that can be administered to a group of subjects simultaneously by means of printed blanks. If motor factors are significant, group administration will seldom be possible.

SENSATION AND PERCEPTION

Example 8. The ordinary chart used by opticians with groups of letters of varying size gives a rough measure of visual acuity. The smallest letters that the subject can read at a distance of 20 feet indicates his acuity. For finer measurements letters of constant size are placed at such a distance that they are illegible and are moved toward the subject until he can read them. This maximum legible distance is noted. Frequently a single symbol, such as the letter **E**, or a circle with a small break in the circumference is used; this is turned with the opening pointed in various directions and the maximum distance is found at which the subject can correctly state the direction. These latter methods have the advantage that they are less subject to coaching. The writer

knows a very nearsighted person who passed a routine examination for a position by purchasing all the different optical charts in the city and memorizing them so that when he saw the large letter at the top (which he could barely read at 20 feet) he could recite the remaining invisible letters. In this particular job the use of glasses was little handicap, but in an employment office there might be situations where such deception would be disastrous.

Example 9. Devices are on the market which measure not only visual acuity but also a number of other variables by means of a stereoscope and ingeniously devised slides. Small crosses, for example, appear in the field with digits of various sizes in the center and the subject indicates the highest number he can read. By putting the digits now on one side and now on the other side of the stereoscope slide, acuity of the eyes can be determined separately. The apparatus is also useful in many other determinations such as the coordination of the two eyes in binocular vision.

Example 10. Color blindness may be measured roughly by a booklet in which each test page is made up of a lot of small dots of varying hue. They are so arranged that for the normal eye there will be an obvious pattern on the page, usually one or two digits. The color-blind person, however, will fail to see these numbers. If the subject has difficulty in distinguishing red from green and the digits consist of red dots and the background green dots, they will all look alike to him and he will not see the numbers, whereas the normal eye immediately sees the red numbers. Other pages are arranged so that only a color-blind person can see the digits; the normal one cannot. The test does not give any refined measurement but is widely used for rough diagnosis. More elaborate apparatus involving spectral light with accurately controlled intensity is necessary for careful determination of defects in color vision.

Example 11. Auditory acuity is most effectively measured by an audiometer. With this complicated electrical device it is possible by tuning a dial to produce from an oscillator sounds of a wide range of frequencies and intensities. A crude but fairly satisfactory test uses a small steel ball placed at a constant distance from the subject's ear and dropped a variable distance upon a metal plate. Miniature pliers hold the ball and are released by

slight pressure. The apparatus can be calibrated in terms of the minimum distance of fall that is audible.

Example 12. Kinaesthetic (muscle sense) discrimination may be checked by having the subject press down on a spring scale such as a postal balance until the indicator as seen by the examiner reaches a certain point. The subject then is required to reproduce this pressure by remembering it kinaesthetically and his error is recorded. The same procedure can be followed with a dynamometer which the subject grips with a force sufficient to push the indicator to a designated point.

Example 13. To measure kinaesthetic perception involved in turning the wrist, an apparatus, consisting of a dial, an indicator, and a rotating horizontal handle, may be used. One end of the handle extends through a screen and is then bent at right angles; as the subject turns the horizontal portion, the other end of the handle moves along an arc on the other side of the screen. The indicator, visible only to the examiner, is upright at the beginning of the test; the subject moves the handle clockwise until it comes to a stop. The subject, attempting to remember how it felt in this position, tries to reproduce the motion he has just made after the handle is turned upright again. Prior to this second clockwise movement, the stop is removed by the examiner. The error can be noted in degrees on the arc along which the indicator moves.

Example 14. Speed of reading may be investigated by a series of short paragraphs in each of which there is one wrong word. The following is typical:

Frank had been expecting a letter from his brother for several days. As soon as he found it on the kitchen table, he ate it as quickly as possible.

The subject goes through the paragraphs marking the wrong words as quickly as he can. Obviously he must read the paragraph in order to locate the wrong word which is usually toward the end of the paragraph.

ATTENTION

Example 15. A significant aspect of attention is its range, that is, the number of discrete impressions to which a subject can attend simultaneously. Some textile operators, for example, have

to watch several machines at once. Some form of short-exposure apparatus is necessary for such a test. One type comprises a shutter containing a slit which is pulled by a spring across the exposure field exactly like the focal plane shutter in some cameras. The card containing the material to be presented is placed in a rack behind the shutter and as the slit moves across the field the material is exposed for a fraction of a second. The speed of exposure can be regulated by the width of the slit and the tension on the spring. Again, an actual camera shutter (compur type) may be used to expose the field. The subject may look through this directly, or a small projection lantern may have the camera shutter attached in front of the lens. At any rate, given some device like this which affords a constant but brief exposure, the range of attention may be measured by presenting different numbers of stimuli—for example, three capital letters, then four, then five. The number is increased until the subject reaches his limit and is unable to reproduce them after the single exposure.

Example 16. A more frequently measured aspect of attention is the ability to concentrate or to operate at a high level of attention for a period of time. Typical tests for this capacity involve some type of cancellation. The subject may be given a page of disconnected letters and required to cancel every *a* on the page. If greater complication is desired, a page of random numbers may be provided; the subject underlines every pair of adjacent digits whose sum is 10. Or he may cross out every 2 and draw a ring around every 3 until he comes to a 7 and then cross 3 and ring 2 until he comes to another 7, and so on.

Example 17. Another test of this order has been used quite extensively with clerical workers. Each item involves a pair of numbers ranging from 3 to 12 digits which may be identical or may be slightly different, such as: 3, 6, 8, 5, 9, 2—3, 6, 8, 5, 8, 2. The subject goes rapidly through a page of these pairs, indicating by appropriate check marks whether the two numbers are the same or different.

Example 18. A still different aspect of attention may be measured by the following test in which the subjects find the consecutive numbers in order, that is, find 11 then 12, and so forth. If the examiner does not wish to watch the subject in order to insure that he actually does locate them in order, the subject may be required to mark *a* after 11, *b* after 12, and so on. If he does not

take the numbers in order, he is apt to become confused and write the wrong letters.

26	52	39	24	53	37	16
14	33	18	47	12	21	56
49	44	59	29	55	31	42
35	20	11	50	15	46	27
58	41	28	38	57	34	48
30	23	54	45	19	13	40
17	32	36	25	51	43	22

LEARNING

Example 19. One of the conventional tests for briefly determining ability to learn or to form a new set of associations involves the substitution of symbols from a code for a series of numbers. The following is typical:

[illegible]

The subject writes under each number the corresponding symbol from the code at the top of the page as shown by the first few items. Enough blank numbers are provided to occupy the subject for the desired length of time. At the outset, of course, reference is made to the code for every number, but a subject who learns readily will soon remember some of the symbols without referring to the code and hence will work more rapidly and make a higher score. Various codes with other symbols are possible; if greater complication is desired, the entire alphabet may be used and a symbol substituted for each letter.

Example 20. Another type of learning test which involves more actual coordination employs a maze constructed of sheet metal with grooves in it. The subject traces a stylus along the grooves from the starting point to the finish, attempting to eliminate blind alleys as quickly as possible. In this process his hand is concealed so that the task is performed without the aid of vision.

Example 21. Maze learning may also be tested by a printed plan or diagram of the maze on which the subject traces with a pencil the correct pathway. The following maze is typical, and may be set up on a typewriter.

```

ACCCXXCXCXCCCCCCCCC
CXXXCXCXXXXXCXXXXC
CCCCCCCCXCXCCCCXCX
CXXXCXXXXCXXXXXXXXXC
CXXXCXXXXCXXXXXXXXXC
CXXXXXXXXXCXCCCCCCC
CCXCCXCXXXCXXXXXC
CXXCCCCCCCCXCXCCCC
CXXCXXXXXXXXXCXXXXC
XXXCXCCCCCCCCCXXXXC
CCCCXXCXXXXCXXXXC
CXXXXXCXCCXCXXXXC
CXXXXXCXCCXCXXXXC
CXXCCCCXCXCCCCXC
CXXCXXXCXCCCCCCCC
CXXCXXXCXXXXXXX
CCCCCCCCXCXXXXCCCC
CXXCXXXCXXXXCXXXXC
CXXCCXXCCCCCCCCXXZ

```

The subject starts at A and traces a continuous line to Z, keeping on the letter C and always moving the pencil sideways or up and down, i.e., never moving diagonally. If tests of the maze type are repeated, the improvement gives some indication of learning ability.

Example 22. A rational learning test numbers the letters A to J inclusive in a random order from 1 to 10, the number of each letter being unknown to the subject. The examiner calls out the first letter, A, and the subject guesses numbers until he guesses the correct one for A. The examiner then informs him that it is correct and calls out the next letter, B, whereupon the subject again guesses until he reaches the correct one. This process is continued until the subject has guessed all the numbers correctly twice in succession.

ASSOCIATION

Example 23. While free association tests are occasionally used in which the subject is given a stimulus word and then speaks or

writes words as rapidly as he can think of them, it is usually desirable to control the association process in some way, such as having the subject give or select synonyms or opposites of certain words.

1. RETURN is the opposite of ADVANCE; SURROUND; RESOLVE; GO.
2. ACTIVE is the opposite of PERSON; PASSIVE; NEUTRAL; DESPONDENT.
3. CONVEY is the same as CONDUCT; TRANSPORT; LIFT; GUIDE.
4. OPERATE is the same as REFINES; DISTILL; SURGEON; MANAGE.
5. CHARITABLE is the opposite of UNTRUE; ACT; MISERLY; UN-FRIENDLY.

etc.

The subject underlines that one of the four alternatives which correctly finishes the sentence. The first two lines are correctly marked.

Example 24. Another widely used test that may perhaps be classed here involves analogies.

1. Gun: shoots:: knife: RUN; CUTS; HAT; BIRD.
2. Handle: hammer:: knob: KEY; ROOM; SHUT; DOOR.
3. Camp: safe:: battle: WIN; DANGEROUS; FIELD; FIGHT.
4. Egg: bird:: seed: GROW; CRACK; PLANT; GERMINATE.
5. Cloud-burst: shower:: gale: BATH; BREEZE; DESTROY; WEST.

etc.

The subject underlines that one of the four alternatives which is related to the third word in the line as the second word is related to the first.

Example 25. Another kind of association test is quite common but not so applicable to industrial problems except in cases of suspected psychopathic tendencies. The test consists of giving to the subject one at a time 100 different words to each of which he replies with the first association that occurs to him. Data are available on the frequencies with which various responses are given. Pronounced deviations from normal responses arouse the suspicions of a clinical psychologist. For example, the first word in the list is "table" and 27 per cent of the subjects give the word "chair." If a person responds with the word "cycloid," it is distinctly atypical. While little significance would be attached to a single peculiarity, a considerable number of atypical responses may be diagnostic. As mentioned previously, however, this asso-

ciation test is more useful in clinical practice than in routine employment work.

MEMORY

Example 26. Memory Span.

8	5	7	3						
9	4	2	1	5					
7	3	2	6	4	9				
2	9	5	3	8	7	1			
5	9	7	4	8	6	1	3		
8	3	1	5	7	4	9	2	6	

The first row of numbers is read aloud at the rate of one digit per second. The subject listens during the reading and then immediately writes the numbers from memory. This procedure is repeated with the next row (five digits), then with six digits, seven digits, etc. The subject's score is the maximum number of digits that he can reproduce after one presentation. Several lists like the above are, of course, used. The method may be varied by having the numbers printed, each row on a separate card, and showing them to the subject for a length of time sufficient to allow about one second for reading each digit. This involves visual rather than auditory memory span.

Example 27.

book	shelf
garden	spade
letter	stamp
watch	time
rain	umbrella

etc.

The examiner reads the pairs of words rhythmically. A metronome sounds one beat a second and the examiner reads "book" on the first beat, "shelf" on the second beat, pauses on the third beat, reads "garden" on the fourth, "spade" on the fifth, and "letter" on the seventh, etc. This serves to group together the two words of each pair and the subject is required to remember these two together. As soon as the list of perhaps twenty pairs has been read, the examiner gives the first word of each pair and the subject writes down the one that went with it. For instance, the

examiner says "book" and the subject writes "shelf" if he can recall it. Or the subject may be provided with a blank containing all the first words of each pair and be given a certain time to write down all the second words he can recall. He may even have a blank of this form:

book page; shelf; title; case.

garden flower; lawn; spade; plant.

and be required to check that one of the four alternatives which was previously presented with the first word on the line. The same sort of experiment may be conducted visually, the pairs being shown in succession at a small window in a specially constructed apparatus. They may be typed on adding-machine ribbon, each pair on a separate line, and fed along in guides behind a slit in the apparatus by pulling the ribbon. This usually necessitates individual examining, but it is possible to place this device in the exposure field of a projecting lantern such as a Balopticon and throw the words on a screen at the front of the room in which a group of subjects is sitting.

REACTION TIME

Example 28. The measurement of simple reaction time involves presenting a visual or auditory stimulus such as the disappearance of a light or a metallic click, whereupon the subject reacts with a telegraph key and the time is recorded. For refined laboratory studies it is customary to employ some type of chronoscope which has a very constant speed motor controlled by a tuning fork and which measures in thousandths of a second. For industrial use it is generally adequate to use a less expensive commercial device which records in hundredths of a second. It involves a synchronous motor (most alternating current supplies are very constant in frequency now that we have electric clocks) and a magnetic clutch so that when a circuit is closed a hand in front of the dial of the chronoscope starts moving and when the circuit is broken it stops. A small rectifier to operate the magnetic clutch is embodied in the timing device. The dial is calibrated in hundredths of a second and may be set back to zero after each reaction. The circuits are a bit complicated, usually involving some relays; they need not be described here.

For auditory reaction a telegraph sounder may be used in which a metal armature comes against a stop, thus actually closing an electric circuit which starts the chronoscope. For visual presentation there may be a small point of light such as a diaphragm in front of a lamp which disappears. It is somewhat preferable to have the subject's reaction consist of opening the telegraph key rather than closing it. At the warning "ready" a few seconds before the stimulus, he grasps the key and closes it, either pressing down on it on the table or, if it is suspended, grasping it between thumb and finger. Immediately upon his perception of the stimulus, he releases the key. A considerable number of reactions are measured to strike an average and it may be desirable also to determine the variability of reaction time, that is, whether the subject always reacts consistently near the average or is sometimes very slow and sometimes very fast.

Example 29. In choice reaction time the subject has to choose between certain alternatives and react accordingly. He may have two lights spatially separated and two telegraph keys. If the right light disappears, he operates the right key and for the left light the left key. The circuit can be arranged so that an incorrect response produces a sound in a telephone headset for the examiner. It may be desirable to have these visual stimuli in the periphery of the field. In this case the subject's head is fastened in a headrest at a constant distance from the box containing the stimuli, and he fixates a point between them in order to keep constant the angular displacement from the fovea.

Example 30. Reaction time may be measured by using other muscles than the fingers. The subject may stand on a pedal and react by stepping off in response to the stimulus. This test involves not merely quickness of reaction but also ability to move the bodily weight quickly. Reaction time has been measured with a small traffic light as a stimulus and with automobile controls as the reaction mechanism. The subject, at the appearance of the red light, moves his foot from the accelerator to the brake pedal.

In measuring visual reaction time, it is important to take precautions so that the subject cannot receive any auditory cue, because auditory reaction actually is quicker than visual. If all the apparatus is in the same room and there is a click in a relay, for

example, at the time the visual stimulus is given, the subject may react to this click rather than the light.

SPACE PERCEPTION

Example 31. If a picture of a rhomboid-shaped card with two holes punched in it near adjacent corners is shown in two different positions, it is rather difficult to tell whether one is looking at the same or different sides of the card. On the test blank many pairs of this sort are provided, and the subject checks each pair to indicate whether it represents the same or different sides of the card. A somewhat similar test involves pictures of a human hand in various unusual positions, the subject in each case indicating whether it is the right or the left hand.

Example 32. The "wiggly block" test may be classed with space perception tests although it may measure other things as well. It consists of a block of wood that has been sawed into nine wavy pieces. Two cuts are made through it in one general direction; the cuts are roughly parallel but each one has three pronounced waves in it. Then two more cuts are made roughly at right angles to the first, likewise with three waves in each. This reduces the block to nine irregularly shaped pieces which must be assembled as quickly as possible.

REASONING

Example 33. A series of arguments like the following is given and the subject marks an item *X* if the conclusion is true and *O* if the conclusion is false.

- ..X.. 1. John's birthday is after Harry's and Harry's birthday is after Tom's. Therefore Tom's birthday is before John's.
- ..O.. 2. William has a brother George who has a son, Henry. Therefore Henry is William's uncle.
- 3. Silver is heavier than iron. Copper is lighter than silver. Therefore copper is heavier than iron.
- 4. Jones owes Smith one hundred dollars. Brown owes Jones one hundred dollars. The two debts will be settled if Smith pays one hundred dollars to Brown.
- 5. All members of the Country Club are members of the Polo Club. Smith is not a member of the Polo Club. Therefore he is not a member of the Country Club.

Example 34.

1. AOUUA UOAU AOAA	2. AAOUAU OUUAAA AAAAOU	3. AAUOUA UAAO UUUOAU
4. UAUUUAOA UUAUUAOU AUAOUUA	5. AUUAOAAAU AOU AUOAA	6. AOUUU UUAAO UUUUAO

etc.

The letter *O* in each line bears a certain relation to the rest of the line. The same relation holds for all three lines of the given problem. For instance, in the first problem the *O* is "second from the left" in all three lines. In problem 2 the letter *O* occurs "before the first *U*" in all three lines. In problem 3 the answer is "fourth from the left"; in problem 4 it is "after the second *A*." The subject writes these phrases under each problem.

SPEED OF DECISION

Example 35.

OOAEUAUAA	OEAOUEOOUA	EAOEUEOAOE	UUEOUAOEOE
OAUAUAOEA	UOUACOAEOU	EUUEAAOEOA	OAUAEOUEUA
AUAAEAEOAE	OUOEEOAOAO	OEUAEUEAUU	UOEAEUAUOU
UUAOEAAAUU	EOEAOUAUOU	EAOEAOEOUE	OAUEAAUOEU
AUAEEAEEA	OEOEOUAOO	EAEUUOEUEU	EOAUUOEUAU
(A)	(O)	(E)	()
EUOAEEOUAE	AEOUUAEOO	UAUEUAUEUA	AOEUOAEUOE
OEAEEUEAEE	OUEOEOOUO	AOUAUAUOA	OEOUOUOUO
EAEEOEUUEU	AOEAOOUOA	OUUOUUOEUE	AUOEAEAOAU
AEOUEEOEOE	OUAOAEEOEO	UEAUAEUOE	OAEOOAEAO
EAEEOEUAE	OUOEOAOOUO	EUOEUEEUUO	EOAUAEUOE
()	()	()	()
EAUEOAUOE	AEAOAEOAUA		
UEOEOAOOE	OAUAUAUAEU		
EUEAOUEAOE	AEAAEAOA		
EAUEEAUOE	OAEUAOEUA		
AEUAUEUAOE	AOAEUAOAU		
()	()		

etc.

The subject is allowed five seconds to glance at each square and determine which letter predominates. The result of this quick de-

cision is written in the brackets below the square. A typical blank comprises 48 squares of this sort. The examiner gives the signal "Begin" and in five seconds says "Mark." Thereupon the subject immediately writes his judgment in the first bracket and looks at the next square. Five seconds later the examiner again says "Mark" and the subject immediately writes under the second square and turns to the third. After the examiner has said "Mark" 48 times the subject is prevented from writing further, so that if he has not kept up with the examiner there are some unmarked squares to reveal this fact.

INGENUITY

Example 36. Various tests of the puzzle type may be standardized as measures of ingenuity.

Animals and Birds		Fruits and Vegetables	
eehps	beeelt	aelpp	inprtu
ekmnoy	binor	aaabnn	acenp
aberz	eginop	aegpr	amoott
ehnort	kknsu	alntuw	abens
aekns	aeelsw	elmno	acorrt
		etc.	

The letters of each word are arranged alphabetically rather than in the normal order. Those in the first group are names of animals or birds; those in the second group, fruits or vegetables. The subject determines what word the letters would make if put in the correct order and writes it after the corresponding letters. He is given a short time limit for each group of words and in that time skips around and gets as many of that group as possible. Other categories such as proper names, furniture, and cities may be used.

Example 37.

1. <u>s</u> pot	<u>m</u> ind	long	
2. <u>b</u> all	<u>m</u> eat	<u>s</u> and	four
3. <u>s</u> ift	<u>p</u> lay	arm <u>y</u>	
4. tw <u>i</u> g	hope	fill	flag
5. hand	note	grab	
		etc.	

In each problem, if one letter is taken from the first word, one

letter from the second, and another letter from the third, and they are put together in that order, they will form the name of an animal. In the first line the three underlined letters spell *pig*. In the second line the answer is *bear*.

Example 38. The following is one type of "completion" test.

Ann t... the poker and began breaking the big l... of coal in the g.... as she said this. Little spirals of greenish yellow s.... escaped from the cracks made by the p.... then jetted into f.... She was so s.... for this woman before her that she l.... doggedly at a lump of coal a... the w.... that she was speaking.
etc.

The subject fills in the blanks in the text. The test may be varied by having the number of missing letters indicated by dots or other symbols, as in the present instance, or by giving no clue to the length of the word. The initial letter may or may not be given.

ABILITY TO FOLLOW DIRECTIONS

Example 39.

If the word contains the letters E, A, and R, mark it 1.
If the word contains the letter E but not A and R, mark it 2.
If the word contains the letter A but not E and R, mark it 3.
If the word contains the letter R but not E and A, mark it 4.
years..... reason..... taint..... addition.....
height..... action..... beguile..... island.....
alfalfa..... bright..... bureau..... office.....
right..... print..... rough..... when.....
verbal..... rocky..... lurid..... forbear.....
etc.

The test can be made more complicated by using other combinations such as "E and A but not R"; "A and R but not E." In general, the more combinations involved, the more difficult will it be to follow the directions.

Example 40.

Make a cross here...and a circle here...and cross out the second and third letters of this word—PECULIAR. If you think there was a war in 1942 put in a number to complete the sentence: "A horse has ... feet." If Tuesday comes after Monday make two crosses here ... but if not make a circle here ... If it snows hardest in summer make

a cross here . . . but if not pass on to the next question and tell what you wear on your hands in cold weather Draw a line between the names of these two boys George Henry and write "no" if 2 times 3 is 6. Notice these five letters A B C D E. Draw a line from A to D that will pass above B and below C. Notice these numbers—3, 5. If a rock is heavier than a feather write the larger number here . . . but if not write the smaller number. . . . Give a wrong answer to the question "How many days in a week?" . . . If sand is good to eat write "no" here but if it is not, write "yes" here If fishes live in the water make a triangle here and a square here Cross out every letter E in the words between triangle and the square which you just drew.

etc.

TESTS OF INTELLIGENCE

We have thus far been illustrating tests of special capacity. There are many industrial situations where a person is being considered for a job that requires unusually good mental equipment along a particular line. He may not need all-round ability, but rather something quite specialized for the particular limited group of operations he is to perform. The tests thus far described are designed largely to meet this situation. On the other hand, as suggested earlier, there are situations in which no outstanding special capacity of this sort seems necessary. The individual needs to be generally alert, perhaps, and able to make moderate adjustments to the conditions of his job, but he does not need quick reaction time any more than he needs speed of association or ability to judge distances. Such a person is usually said to be at a certain intellectual level or to possess a certain degree of intelligence.

Nature of Intelligence. This is not the place for an elaborate discussion of the nature of intelligence, for, just as with the tests of special capacity, the crucial point is whether the particular tests facilitate vocational prediction, regardless of what, in the last analysis, they measure. Electricity could be measured and used to ring doorbells or operate telegraph tickers before it was defined or its exact nature known. Similarly, the psychologist can measure intelligence for practical purposes even though he is not certain what it is. Scientists' conceptions of intelligence appear to depend somewhat on the interest with which they approach

the problem. A statistician is quite apt to think of it as a general factor which causes intercorrelations between miscellaneous mental tests; i.e., people who make high scores in one kind of mental test often make somewhat similar scores in a good many different kinds of tests, and vice versa. A person with biological interests may be inclined to conceive of it in terms of ability to adapt oneself to one's environment and make the appropriate adjustments as the environment changes. A certain degree of flexibility in behavior seems demanded by human society. One with a physiological trend is inclined to think of intelligence with reference to neural aspects—the degree of plasticity of the nervous system and facility in forming new connections and patterns therein. Investigations have been made of the correlation between intelligence and reaction time. To the business man it connotes mental alertness, ability to follow instructions, to analyze a situation, to learn readily, and to “catch on” to new situations. This last conception comes nearest to that which would be adopted by a personnel psychologist if it were necessary for him to commit himself as to the nature of intelligence. There appears to be some general capacity that gives one a better chance for survival in the economic struggle. One man can start out in any one of a dozen lines of work and be successful in any of them. Whatever line he enters he gets a good start, he learns his duties rapidly and accurately. Another man, though he may try many types of vocation, is practically doomed to failure in any of them. He is “dumb”; he does not “get the idea”; he cannot adjust himself; he is slow, and he often does the wrong thing or at least fails to do the right thing. This is the type that floats around trying one job after another, losing it, and frequently becoming delinquent. As far as employment psychology is concerned, we may say that the first man has high intelligence and the second has low intelligence.

Individual Intelligence Tests. Intelligence tests have been devised for both individual and group administration. The latter procedure is preferable in most employment situations because of time-saving. The former, however, is important when the examination is somewhat clinical in nature, as for instance in investigating problem employees. An experienced examiner will obtain more from the examination than the mere test score. He

observes the way the subject goes at the items, any peculiar associations and any emotional manifestations. Both individual and group tests will be illustrated.

Example 41. The most widely used individual intelligence test is the Binet-Simon test originally developed in France and subsequently standardized by various people in this country; the most widely used standardization is that by Terman [16]. The test comprises a series of questions or tasks which the average child of each age can answer or perform correctly. If, for example, a person passes the 5-year-old items but fails on the 6-year-old he is accorded a *mental* age of 5 regardless of his *chronological* age. There is the possibility of failing an item at one age and passing some at a higher age. The scoring takes account of this fact by allowing certain months' credit for each item. The point is that the average child whose chronological age is five will test exactly at five or have a mental age of 5. Consequently anyone can be given the test and his mental age determined. If he passes only the 11-year-old items although he is an adult, he is given a mental age of 11. The present form of the Binet test ranges from a mental age of 2 years to 22 years and 10 months. The reason for not extending the range to, say, a mental age of 50 years will be discussed below. Typical items from the Binet test are given to show the kind of things called for at a few ages. The 2-year-old level includes the following:

1. A toy cat is placed under one of three boxes, then a screen held in front of it for three seconds and the child is required to "find the kitty."
2. Several objects must be identified as they are presented on a card. For example, "Show me the ball", "Show me the scissors."
3. On a paper doll the child is asked to point to the eyes, the feet, the nose.
4. From a simple form board, circular, square, and triangular blocks are removed while the child is watching and he is required to put them back in the right holes.

Some of the items at the 9-year-old level are as follows:

1. Look at a card with a very complicated diagram for ten seconds and reproduce it from memory.
2. Rearrange groups of words to make a sentence. For example: "A have dog I found" or "Wool the was coat of made."

3. Recognize absurdities, that is, "what is foolish" in sentences. For example: "A father wrote to his son, 'I enclose ten dollars, if you do not receive this letter, please send me a telegram.'" or: "A soldier complained that everyone else was out of step except himself."
4. State similarities and differences between certain items such as honey and glue, a pencil and a pen, a banana and a lemon.
5. Give rhymes as rapidly as possible for 30 seconds with such words as "date" or "head."

At the average adult level there are the following:

1. Give the meaning of abstract words such as generosity, attendance, envy, authority.
2. Suggest a solution for problems such as the following: Going to the river to bring back exactly two pints of water when one has only a five-pint and a three-pint container.
3. Complete analogies like the following: "A rabbit is timid. A lion is _____. " "Trees are terrestrial. Stars are _____. " "A group made up of dissimilar things is heterogeneous. One made up of things which are alike is _____. "
4. Discover the system in a simple code and apply it in writing another word.
5. Interpret proverbs such as "We only know the importance of water when the well is dry."
6. Show orientation in problems like the following: "Suppose you were going south and you turned left and then right, what direction are you going now?"

Group Tests. Historically the most significant group intelligence test was the Army Alpha devised for military use in 1917. It is based on the principle that a considerable number of special capacity tests may be combined to give a fair indication of general capacity or intelligence. This test served as the prototype for a great many of the group intelligence tests that have been developed subsequently.

Example 42. The following are excerpts from such a test pitched at approximately high school level but presumably useful for many types of industrial workers. It is an *omnibus* test. Instead of grouping the various kinds of test items so that a page of one type is completed before the next is turned to, as in the original Army Alpha, the different types are intermixed. The subject does a very few items of one sort, then a very few of another sort, etc.

He may even do only one item of a given kind at a time, thus shifting rapidly from one kind to another. Four types of multiple choice items in which the subject selects the correct alternative rotate through the present test:

1. What is the opposite of success?—HAPPINESS; FAILURE; HONOR; DEPRESSION; JOY.
2. In what war was Gettysburg famous?—WORLD WAR; 1812; SPANISH-AMERICAN; CIVIL; MEXICAN.
3. If you go camping and pay \$1.00 which is one-fourth of the total expense, what was the total cost of the trip?—\$2.00; \$3.00; \$5.00; \$3.50; \$4.00.
4. Why are airplanes used for mail service?—THEY ARE SAFE; THEY CANNOT BE ROBBED; THEY ARE NOT STOPPED BY STORMS; THEY ARE CHEAPER THAN TRAINS; THEY ARE VERY FAST.

Example 43. Another test of this sort at the college level and possibly useful for executives is that developed at Ohio State University. Some items involve fairly difficult opposites such as the following:

1. IMPROMPTU is the same as PROMPT; EXTEMPORANEOUS; IMPROPER; SPEECH; BRIEF.
2. HUMBLE is the opposite of HAUGHTY; UNASSUMING; OBVIOUS; GUILELESS; PLAINTIVE.
3. ACCENTUATE is the same as ATTENUATE; DRESS; ACCENTUATES; UNNOTICED; EMPHASIZE.
4. FRIVOLOUS is the opposite of COY; TRIFLING; SERIOUS; SAGACIOUS; RIGHTEOUS.

Another type of item involves two words that are related, followed by a third, and then alternatives from which may be selected the one that bears the same relation to the third word. For example:

1. Boy : boys :: Man : GIRL; MEN; MAN'S; MEN'S; GENTLEMEN.
2. My : I :: His : ITS; HE; ME; HIM; HIS.
3. Communicable: communication :: disruptive : DISRUPTION; DISRUPTIBLE; DISRUPTABLE; DISRUPT.

The next kind of item involves reading a complicated passage and answering questions about it. For instance:

The ordinary form of mercury thermometer is used for temperatures

ranging from -40° F. to 500° F. For measuring temperatures below -40° F. thermometers filled with alcohol are used. These, however, are not satisfactory for use at high temperatures. When the mercury thermometer is used for temperatures above 500° F. the space above the mercury is filled with some inert gas, usually nitrogen or carbon dioxide, placed in the thermometer under pressure.

The paragraph continues and then is followed by questions like the following:

What chiefly determines the upper temperature at which thermometers can be used? (1) weight of mercury, (2) gas pressure, (3) melting point of glass, (4) amount of gas, (5) rates of expansion.

What word meaning "inactive" is used in the paragraph? (1) limit, (2) inert, (3) limited, (4) not satisfactory, (5) ordinary.

Example 44. Performance Tests. Among the non-verbal or performance tests, one of the common types is the *form board*. This test appears in many varieties. For instance, a board is provided, from which holes of various shapes have been cut—square, circle, cross, star, diamond. Blocks are provided of the proper shape to fit these holes. The subject's problem is to fit the blocks into the holes as quickly as possible, and a record may be taken of the moves he makes and of the time. This particular board is perhaps too simple for ordinary industrial use, but the principle can be extended to include all degrees of difficulty. For example, there may be a single rectangular hole and a number of rectangular and triangular blocks which if fitted together in the proper manner will fill the hole. Numerous other complicated patterns have been devised.

Example 45. It is possible to arrange a form board test so that it can be administered to a group of subjects by means of printed blanks. For instance, if a rectangular hole in the board is to be filled completely by two rectangles of different sizes and two triangles, a picture of a blank rectangle of the appropriate size and proportions can be presented, and at one side a picture of the four small pieces drawn to scale which if properly placed will fill the larger rectangle. The subject merely sketches in the blank rectangle to show how these four parts would fit. Items of this sort presumably measure the same mental characteristics

as when the subject actually manipulates the blocks and puts them in the holes.

Example 46. Other performance tests similar to puzzles have been standardized, such as a picture cut into a number of irregular pieces which the subject must fit together. A picture completion test involves a picture showing a lot of miscellaneous activities in progress; small half-inch squares are removed at various places in the picture. The subject is provided with a lot of small squares, each with a picture on it, and he must pick out appropriate ones and fit them into the holes in the larger picture. If a person is shown standing on a ladder in an apple tree with his arm extended, a basket of apples on the ground below, and a square cut out between the hand and the basket, it is obvious that an apple is being dropped. If the subject, however, locates the picture of a shoe and inserts it at this point, he fails to see the relationship.

Kinds of Intelligence. The preceding discussion has dealt with what is often designated as abstract or verbal intelligence. This is the kind most frequently involved in typical intelligence tests. The subject is presented with problems that have an abstract ideational content. But it is generally agreed that there is something which may be termed *mechanical intelligence*. Some persons who do not manifest a high general capacity for handling abstract concepts may nevertheless have a distinct general superiority over their fellows when it is a question of manipulating concrete objects. Dealing with things that you can take in your hands and place in different positions and put together in various ways is somewhat different from dealing with words which are mere symbols. By this mechanical intelligence is meant not mere manual dexterity or ability to perform a single mechanical operation, but rather something of a more general character. Just as high intelligence of the abstract type enables a man to be successful in any one of many vocational pursuits that involve this kind of intelligence, so a person with high intelligence of the mechanical type will presumably be successful in any one of many vocations where he deals with concrete rather than with abstract things and manipulates objects other than a pencil. This notion of mechanical intelligence is not as firmly grounded as the other and less actual experimental work has been done in this field, but

it is an aspect of which the employment psychologist should take account in certain practical situations.

There is still another type of intelligence with reference to which less has been done; this is *social intelligence*. There may be instances where consideration should be given a person's general capacity not for dealing with abstract concepts or for handling concrete but inanimate things, but rather for dealing with social situations and reacting to other people. Tests of this type are still in the experimental stage, but if they are subsequently perfected they may be of considerable practical significance for the types of vocation in which social contacts play a large role. A few examples of each of these types will be given.

Example 47. Mechanical Intelligence. One of the most widely used individual tests for mechanical aptitude or general mechanical intelligence consists of assembling a number of small appliances. A box is provided with several compartments. The first contains the three parts of a small simple monkey wrench. The subject is required to put the wrench together by putting the head through the end of the handle and inserting a thumbscrew at the proper place. As soon as he finishes this compartment, he turns to the next in which there are six links of a light chain. These likewise must be assembled in correct fashion. The next compartment contains the parts of a spring paper clip. Other compartments contain a bicycle bell, a coin holder, a spring clothespin, a shut-off for a rubber hose, a push button, a simple lock, and a mouse trap. The subject assembles the items one after the other; he usually has a time limit for the entire examination. The different items are scored according to a special scale. For instance, a perfect assembly of the wrench gives 10 points; if the nut is in the wrong place, 4 to 6 points are allowed; if in addition the head is turned in the wrong direction, only one point is scored. A similar scale is available for each assembled object.

Example 48. The foregoing example is essentially an individual test. If duplicate sets of material are provided, it may be given to small groups simultaneously, the subjects solving the problems in order, with a definite time limit for the whole. To give it on a large scale as a group test involves a considerable outlay. Efforts have been made accordingly to measure somewhat the same factors by means of a printed blank that may be employed like

the usual group test. One such test involves small pictures of a variety of mechanical objects. They are presented in groups of five pictures each. The groups are arranged in pairs and so constituted that each object in one group belongs with an object in the paired group. The objects illustrated in a typical group are as follows:

First Group	Paired Group	
1. screwdriver	A. twist drill	1. ..C..
2. bit stalk	B. anvil	2. ..E..
3. tire pump	C. wood screw	3. ..D..
4. brace	D. tire	4.
5. hammer	E. bit	5.

The objects in the first group are numbered from 1 to 5 and those in the paired group are lettered from A to E. The subject must identify an object in the second group that goes with each object in the first. At the right are the numbers 1, 2, 3, 4, 5. The subject writes after each number the letter of the corresponding object that belongs with it. In the above illustration after 1 he writes C, because the picture of the screwdriver and that of the wood screw belong together. Another set of pictures involves in the first group a valve grinder, spark-plug wrench, throttle, set of coil points, and hydrometer. The paired group contains an accelerator, storage battery, spark plug, engine valve, and spark coil. Other groups involve such things as locks, curtain rods, hinges, telephone construction, gauges, and parts of vehicles.

Example 49. Another test designed for pencil and paper administration involves a series of vertical lines, each with a small gap in it; the subject has to locate these gaps as quickly as possible and trace a continuous line across the page. A series of small circles linked together by straight or curved lines goes through a complicated pattern which the subject has to follow, putting a dot in each circle seriatim. A fairly complicated diagram of straight lines and angles has to be reproduced on cross-section paper, making the angles on the diagram coincide with the intersections on the sheet. Pictures of stacks of small cubes in various patterns have one cube marked and the subject is required to tell how many other cubes are touching it. An ocular pursuit test involves a series of very wavy lines extended from the left of the

page to the right, crossing each other many times. The starting points on the left are designated by numbers; the subject has to follow through with his eyes and put corresponding numbers at the other end of each line on the right side of the page.

Example 50. One of the most comprehensive studies of mechanical capacity was made at Minnesota [15]. It employed some tests of assembling small objects like those described above, also paper form boards and spatial relations tests which were essentially complicated form boards with a lot of small irregular holes and a block to fit each hole.

Example 51. Social Intelligence. The only test of social intelligence published is that by Moss [14]. The following are some of the kinds of items:

1. Judgment in social situations: An employer has two men who do not get along very well with each other. Alternative solutions of the problem are presented and the subject is required to select the best one.

2. Recognition of the mental state of the speaker: "The idea of asking those Baileys." The subject is asked to indicate the mental state of the person making the remark.

3. Observation of human behavior: Statements like the following are to be marked as to whether they are true or false, "We can place little confidence in those who love a lot on slight acquaintance."

4. Memory for names and faces: Photographs are presented, together with fictitious names. Subsequently the photos are shown with only the first names. The subject has to supply the second names.

5. Judgment of facial expression: Photographs are shown of persons registering various emotions which the subject must identify.

Example 52. One other effort to develop a social intelligence test may be described [3].

1. overalls, orchestra, favors, chaperone, program.
2. cheerleader, gown, address, degree, diploma.
3. bride, clergyman, soup, organ, ring.
4. judge, colonel, jury, plaintiff, defendant.
5. coming out, censor, flowers, music, debutante.

Subjects who are more alert to things social will more quickly recognize the social situation in each line and be able to detect the wrong word.

1. "Your lead, partner."—BRIDGE GAME; DANCE; TENNIS; BILLIARDS.
2. "Baby needs a pair of shoes!"—MOTHER TO CHILD; GOLD-DIGGING CHORUS GIRL; GAME OF DICE; TOE DANCER.
3. "Here's how."—TOAST; TEACHER TO CLASS; INTRODUCTION; FAREWELL.
4. "Forty, love."—MAN TALKING TO WIFE; TENNIS MATCH; BROADWAY MUSICAL REVIEW; STOCK MARKET.
5. "Hold that line!"—MANAGER TO SALESMAN; FOOTBALL CROWD; TELEPHONE OPERATOR; FISHERMAN.

The subject is required to indicate in which situation the remark would be most appropriate.

1. "Oh, what's the use!"—DESPAIR; ANGER; GRIEF; PAIN.
2. "Must you do that?"—SURPRISE; ANNOYANCE; IMPATIENCE; FORBEARANCE.
3. "I could die!"—MIRTH; ENNUI; DISAPPROVAL; FEAR.
4. "Attaboy!"—DISCREDIT; PRAISE; SURPRISE; COMMAND.

The subject merely identifies the emotion.

1. You accidentally break a small vase which belongs to your hostess.—CRY; SAY THAT YOU WILL REPLACE IT; GET A DUSTPAN AND CLEAN UP THE MESS; SAY "I AM GLAD BECAUSE I DID NOT LIKE IT ANYWAY."
2. You are dancing and step on your partner's foot.—STOP DANCING; APOLOGIZE; LAUGH; RUN AWAY.
3. Two guests get into a heated argument which is spoiling your party.—INTERRUPT AND TAKE ONE OFF TO DO SOMETHING ELSE; TELL A FUNNY STORY TO THE ENTIRE GATHERING; TACTFULLY CHANGE THE ENTIRE SUBJECT; TELL THEM NOT TO BEHAVE LIKE CHILDREN.
4. You are at a party with others only of your own sex and know no one but the host and hostess.—EAT BY YOURSELF; INTRODUCE YOURSELF TO SOME OF THEM; HUNT UP THE HOST OR HOSTESS; TELL A JOKE.

The subject checks the alternative that he considers the best thing to do.

1. A grand slam is taking:—SIX TRICKS; NO TRICKS; THIRTEEN TRICKS; NINE TRICKS.
2. KDKA is in:—NEW YORK; BOSTON; CHICAGO; PITTSBURGH.
3. Fair catch is used in:—FOOTBALL; BASKETBALL; BASEBALL; TENNIS.
4. The number of beats in a measure for waltz music is:—SIX; TWO; THREE; FOUR.
5. A mallet is used in:—TENNIS; BRIDGE; CROQUET; HOCKEY.

These items deal with everyday information which might be used in small talk. The theory is that a person who is well informed on such items can keep a conversation going more effectively.

Inasmuch as the measurement of social intelligence is still in the experimental stage a few investigations may be mentioned. The test just described was given to university women and the results were correlated with an index of social competence obtained by records available in the office of the Dean of Women. These included participation in student activities, and frequency of teas, bridge games, and dances weighted into a composite score. The various items in the test, weighted optimally, correlated with this criterion of social competence to the extent of .49.

With Moss's test the average scores made by certain occupational groups indicate high scores where they appear plausible. Executives, teachers, high-grade secretaries, and salesmen rate appreciably higher in the test than do low-grade office workers, counter salespeople, and industrial workers [10]. On the other hand, abstract intelligence scores likewise correspond rather plausibly to this occupational hierarchy. (Cf. [17].) When social and abstract intelligence scores are correlated directly, the coefficients are around .50 or a little better. There is also the question as to whether social intelligence is essentially an innate thing such as abstract intelligence is considered to be. Some of the items, like insight into the social situation, obviously require experience. If the test measures something that is actually innate it must be assumed that everybody has had adequate opportunity to pick up information about social things and those who are more alert in this respect would be more inclined to pick up the information. On this hypothesis the test would not be appropriate for some people whose social environment prior to the time of taking the test had been very limited. Thorndike made a factor analysis (cf. p. 107 *infra*) of social intelligence test data and found three factors, of which the principal one seems to be verbal. He concludes that the so-called social intelligence test is "a rather poor test of general intelligence."

Interpreting Intelligence Scores. Some intelligence tests yield a score which consists of a certain number of points. This can then

be standardized in various ways just as in the case of special ability tests. The test may be given to a large number of subjects and the average score computed so that any individual's score can be evaluated by comparison with the average.

For finer standards the percentile method is often used. The individual scores are arranged in order from best to worst. The best one is called the 100 percentile, indicating that the subject equals or exceeds in proficiency 100 per cent of the group. Then a slightly lower number of points is computed, such that those attaining that number of points equal or exceed 99 per cent of the group. This score is called the 99 percentile. Similarly, a 50 percentile individual equals or exceeds half the group. The matter may be made clearer by a brief example. (See Table 10.) Suppose that one person makes a score of 28, another 29, 2 subjects score 30, 3 subjects score 31, etc., up to the best one, who scores 39. In the

TABLE 10. ILLUSTRATING THE PERCENTILE METHOD OF INTERPRETING TEST SCORES

Raw Score	Number of Subjects Making Test Score	Cumulative Number	Percentile Score
28	1	1	2
29	1	2	4
30	2	4	8
31	3	7	14
32	3	10	20
33	6	16	32
34	9	25	50
35	8	33	66
36	7	40	80
37	6	46	92
38	3	49	98
39	1	50	100

third column we see that 2 subjects score 29 or less; 4 subjects score 30 or less, 7 subjects score 31 or less, etc. These last-mentioned figures may now be converted into percentage of the total number of the subjects, namely, 50. These percentages, which appear in the last column, constitute the percentile scores. So, instead of saying that a subject scores 31 points, we may say

that his score is the 14 percentile, meaning thereby that he equals or exceeds 14 per cent of the group in intelligence.

This percentile procedure for conversion of test scores is widely used. In many instances the interest lies in basing standards on a particular group of individuals, such as freshmen in college or office workers or unskilled laborers. The percentile method is a useful way of expressing the standing of an individual relative to the standard group. Furthermore, it makes it possible to compare an individual's standing in one test with his standing in another test. If he is a 75 percentile in one test and a 50 percentile in another, he is obviously superior in the first, although his raw scores (because of the number of test items involved) may not indicate this difference. Incidentally, percentile scores are equally as applicable to special capacity tests as to intelligence tests although less frequently used there.

Some intelligence tests yield not a score in points, but a mental age. This is particularly characteristic of the Binet test above mentioned. Certain questions are given for the three-year level, the four-year level, and so on, and on the basis of the questions a subject answers he is assigned a particular mental age. The usual procedure is then to compute his *intelligence quotient* (I.Q.). This is his mental age divided by his chronological age. If, for instance, his mental age is 12 and his chronological age is 10, his I.Q. is 120;² i.e., he is 20 per cent above the average mentally for persons of his chronological age. If his mental age is 10 years and 3 months and his chronological age 12 years and 9 months, his I.Q. is about 80 (123 months divided by 153 months); i.e., his intelligence is only 80 per cent of what it should be.

When dealing with adults, the above procedure breaks down because it is generally agreed that the type of thing measured by the Binet test stops increasing somewhere in the middle teens and that thereafter a person's actual intelligence does not increase but merely his information or education. Thus a normal adult at the age of 48 might secure a mental age of 16 on a Binet test which would be doing well. Computing the I.Q. directly, however, would give him an I.Q. of 33 which would indicate a low-grade imbecile. Consequently, the standard procedure is to consider the

² It is conventional practice to carry the ratio to two places and drop the decimal.

chronological age 16 for purposes of computing the I.Q. of anyone who is older than this, the assumption being that a mental age of 16 is typical for the average adult. This figure has sometimes been questioned, and the tendency has been to lower it rather than raise it. As a matter of fact, the computations beyond the chronological age of 13 are not made exactly as indicated. Some adjustments are made to take account of the fact that the rate of growth of intelligence is decreasing in the teens. Formulae have been developed to make such adjustment, but it is more convenient to consult tables which have been provided. One can look up the chronological and mental ages and read the I.Q. directly. For example, for both chronological and mental ages of 15 the I.Q. is 105. For a person whose chronological age is 16 or over, the highest possible I.Q. is 115. The I.Q., then, is a useful index of the extent to which the individual's abstract intelligence exceeds or falls short of the average intelligence of people of the same chronological age or, if he is over 16, of other adults.

FACTOR ANALYSIS OF ABILITY

Mention should be made at this point of efforts to apply the technique of factor analysis to determine whether the general ability involved in intelligence tests can be broken down into a limited number of special capacities. We noted earlier that in constructing group intelligence tests the usual procedure is to assemble a considerable number of special tests and note the total score. The detailed technique of factor analysis is beyond the scope of the present discussion, but the results of one investigation may be mentioned [18, 20].

Factor analysis begins with data in the form of intercorrelations between all the subtests under consideration. We have an array or matrix that gives the correlation of the first test with the second, first with third, second with third, and so on. If correlations are high, we assume that the tests overlap to some extent and presumably are measuring the same factor. It is possible that each test actually embodies a certain amount of several factors. If these factors overlap a good deal it may be that with a rather limited number we could explain all the intercorrelations. To put it rather crudely, the first test might involve 50 per cent of factor number one, 30 per cent of factor number two, and 20 per cent of

factor number three. Another test might involve 75 per cent of factor two, and so on.

Earlier statistical studies were directed at finding if there might be a single general factor that runs through all these tests. It was thought that this factor might be what had been called intelligence. More recent work begins without the assumption of a general factor and simply tries to find how many factors would be needed to account for the whole matrix of intercorrelations, and also the actual loading of each factor in each test. It is hoped that the number of basic factors will be much smaller than the number of tests.

The statistical analysis is rather involved but finally results in a series of factor loadings for each test. These figures represent the loadings for the first factor. It is then possible to determine by further formulae how much of the intercorrelation between the variables has been accounted for by this first factor. By extension of the procedure we then determine the loading of the second factor in each test and likewise how much is left over. We proceed until the residual is too small to be worth further analysis. We then know that a certain number of factors will account for practically everything involved in the intercorrelations between the tests. In other words, if each test can be considered as the compound of several factors loaded as indicated, we can see why the tests should overlap and we also see how many factors are necessary to account for the entire matrix.

Up to this point the treatment has been entirely statistical and abstract, with no indication of what the factors are. The next step is to speculate as to the actual nature of the factors by examining the loadings. For instance, if we find that a certain factor has heavy loadings in three tests but is comparatively light in the others, we may consider what those tests have in common. If, for example, they all involve manipulation of numbers, it is possible that this factor deals with number facility. Similarly, if the second factor has heavy loadings in several tests and these tests all require the subject to examine some material and then do something from memory, it is possible that memory is the second factor. By this procedure we obtain some notion as to what the basic factors are. Then if we are interested in improving

or revising the test we may try to devise tests which measure more directly the factors which we think we have discovered.

This technique has been applied by Thurstone to a large number of tests of the sort generally included in intelligence tests [19]. Carrying through the above procedures, he came to the conclusion that seven factors are basic and he designated them as seven "primary abilities." From an inspection of the factor loadings he identified them as follows: 1. Number facility, 2. Word fluency, 3. Visualizing, 4. Memory, 5. Perceptual speed, 6. Induction, 7. Verbal reasoning. He then devised 16 tests aimed directly at these seven primary abilities in order to measure the same factors involved in the original large number of tests.

One other study [21] may be cited which employed 52 tests and carried the analysis to ten factors. The tests, however, included a wider range such as social intelligence, musical ability, and attention. The most important factors appeared to be verbal ability, spatial ability, numerical ability, attention, musical ability, and memory. Several of these coincide with factors mentioned by Thurstone.

Breaking down intelligence into factors like these is of considerable theoretical interest. The employment psychologist is interested too, inasmuch as it may facilitate the construction of better tests or at least tests which in briefer form will be just as satisfactory. If the principal factors are found and tests are reorganized to hit them directly, this may make it possible to save some time and thus facilitate employment programs.

PERSONALITY TESTS

A personnel psychologist often encounters unfortunate situations such as the following. He develops tests for special capacity which correlate satisfactorily with ability in a job and then starts using them to select employees. Many of those selected on the basis of the tests come up to expectations, and do satisfactory work after they have had a chance to learn the job. Others, however, for whom prediction is entirely favorable fail to materialize as good workers. Analysis of individual cases may reveal that this failure is not due to any lack of native capacity. In other words, the tests were all right as far as they went. But perhaps the worker was lazy or reckless or did not get along well with

other people or was "shut-in" and introvert. Industrial psychologists are keenly aware of the need for tests of this aspect of the individual to supplement the capacity tests. It is not merely a matter of what a man *can* do but what he *will* do. This aspect of a person sometimes is evaluated by ratings of acquaintances. These can be standardized so as to possess a fair degree of reliability; they will be discussed in a subsequent chapter. But there have also been efforts to devise personality tests so that these aspects may actually be measured in quantitative terms.

Subjective Tests. Many of the attempts to develop tests for the various aspects of personality have involved what may be called a subjective approach. The subject himself must make statements about himself, such as what he would do in certain situations, whether he is worried about certain things, or what his preferences are for this and that. Tests of this type have proved quite useful in some clinical situations. They do, however, have certain shortcomings for industrial use, as will be pointed out presently. A few efforts along this line may now be described.

Example 53. One of these tests purports to measure ascendancy-submission [1]. It describes situations with several alternative responses, one of which must be checked.

At a lecture or entertainment, if you arrive after the program is begun and find that people are standing but there are plenty of seats available down front, but this would make it necessary to be conspicuous, do you take the front seat: HABITUALLY—OCCASIONALLY—NEVER.

In witnessing a ball game have you intentionally made remarks (witty, encouraging, disparaging, or otherwise) which are clearly audible to others around you: FREQUENTLY—OCCASIONALLY—NEVER.

Example 54. The Bernreuter Personality Schedule is another widely used device [2]. In front of each item are the three symbols "Yes" "No" or "?" one of which the subject checks.

Does it make you uncomfortable to be different or unconventional?

Do you daydream frequently?

Do you usually think things out for yourself rather than getting someone to show you?

Have you ever crossed the street to avoid meeting some person?

Do you ever give money to beggars?

Do you frequently argue over prices with tradesmen?
Do your interests change rapidly?

A scoring key is provided whereby, if neurotic tendency is being scored, certain items on the blank if marked "Yes" are scored plus 3, others plus 2, and some with minus values. Another pattern is available for a score on self-sufficiency; still another for introversion-extroversion, and a fourth for dominance-submission.

Example 55. One other test of this type may be mentioned [9]. It involves a series of items which the subject checks "Yes" or "No."

Is it rare for you to be absolutely sold on an idea?

Do you ever have to fight against bashfulness?

Do you like to have the leisure to sit down and indulge in reverie?

Are you often behind the times in the gossip of the group to which you belong?

Different systems of scoring yield indications of several types of temperament such as schizoid, cycloid, or hysteroid. Such classifications are more frequent in clinical practice, but the test has also been used under industrial conditions.

Example 56. Emotional aspects may be significant in industrial situations and Pressey's X-O Test represents an effort to approach them by a printed test. Some items consist of groups of five words such as "disgust, fear, sex, suspicion, harm"; the subject crosses out any of the five that he considers unpleasant and draws a circle around the one that is the most unpleasant. Other items include five words like the following: "begging, swearing, smoking, flirting, spitting," and the subject indicates which he regards as blameworthy; he also marks the most blameworthy. In items like the following: "injustice, noise, self-consciousness, discouragement, germs" the subject must indicate which are emotionally disturbing and which one is the most so. The fourth type of items includes the following: "blossom: flame, flower, paralyzed, red, sew." The subject marks all the words that are associated in any way with the first word in the line and circles the one that is most closely related to it.

Difficulty with Subjective Tests. Personality tests like the foregoing have been quite widely used in clinical practice and to

some extent in industrial employment. However, they are used in the latter situation with some misgiving. The test necessitates the subject's stating frankly how he feels about certain things. There is the danger that he will on the contrary attempt to answer the questions in a way that will be most favorable to him. There is quite a difference in the attitude of a person coming to a psychological clinic for help and coming to an employment office to get a job. In the former case his attitude is similar to his attitude in a medical office where he describes his symptoms without reservation. If he goes to an employment office and is confronted with such a test he wonders if it will be to his advantage to admit that he has worried about cats or has frequently had to fight shyness. He is inclined to mark the blank in the way that he thinks will be most likely to secure him the job. Hence one should not be misled by the demonstration of the efficacy of the tests in a clinical situation, and assume that they will be equally valuable in solving industrial employment problems. Even in giving capacity tests there is sometimes a possibility of malingerers. For example, a technique has been devised for giving rough intelligence tests in a disguised form. One kind of item in such a test is vocabulary. Rather than being asked the meaning of a certain word, such as "introspective," the subject may be asked in the interview whether his father or his mother was more introspective. He does not suspect that his vocabulary is being checked and his answer will probably show whether or not he is familiar with that particular word.

There are empirical indications that the score in tests of this sort can be "faked" by deliberate effort. A test designed to indicate masculinity was given to some subjects, with instructions to be as masculine or as feminine as possible, and it developed that they could shift their scores "enormously" in either direction [11]. One instance of deliberate alteration of test score by applicants for a job may be mentioned. In this particular case it was desirable to secure salesmen who "knew their way around," that is, who had some contact with socially questionable things and could meet the prospects on almost any level. Consequently, an information test was devised that included items about poker, crap-shooting, chorus girls, and the like, for the purpose of finding out whether the prospective salesmen actually were

familiar with such matters. The intention was to hire those who did show this familiarity. However, a group of applicants who took the test evidently suspected that the results were going to be interpreted in just the opposite fashion; hence practically no one in the group knew the difference between a "full house" and a "flush" or were acquainted with "Little Joe." The test completely misfired.

If the test is given individually by a person with clinical training he may be able to secure sufficient rapport with the subject so that adequate cooperation can be insured. In most employment situations, however, this possibility is questionable. Therefore it is desirable to look toward more objective measures of personality which do not necessitate the subject's committing himself on any self-evaluation. Such techniques, with few exceptions, are scarcely beyond the experimental stage. A few efforts of this sort may be described without implying that such tests are available in final form. As a matter of fact, some employment men have somewhat informal but objective procedures of their own. The writer knows of one organization with a large sales force which at the annual sales convention has a "camp" for a few days. They make it a point to have the men who have recently joined their sales force at this camp, and they "initiate" them. The initiation is rough and includes throwing them into the swimming pool, the idea being to see which ones can't "take it." The young salesman who gets angry and wants to fight somebody is apt to have his connection terminated in the near future. Even initiation into a professional organization which involves untactful remarks about the neophyte while he is attempting to make a speech or carry out directions is rather revealing in the same way.

A number of tests of this sort have been devised by the Character Education Inquiry [7, 12]. The techniques were developed for the most part with children and in some instances would not work with adults, who would be less naïve. Some of the tests dealt with honesty.

Example 57. A series of circles is arranged on the blank with their centers all approximately on a large circle. These circles are of various sizes. The subject places his pencil at a designated point, shuts his eyes, and attempts to make a cross in each circle. He is given several trials, scoring each trial himself before doing

the next. By trying the test on persons who are actually known to have their eyes closed, it can be determined just what are the chances of hitting some of the smaller circles. If the subject does considerably better than this probable expectation, the presumption is that he "peeked." In another variation of this test there are six squares of different sizes, one inside the other, thus affording a continuous pathway between each two squares. The subject starts with his pencil at a designated point in each pathway and with eyes closed traces around through the pathway to the starting point. With the shorter pathways a correct response is possible, but with the longer pathways it is practically impossible. This test can be evaluated in the same manner as the one with the circles.

Example 58. Another test for honesty tempts the subject to make overstatements. He is given a set of preliminary questions such as:

1. Can you swim?
 2. Can you skate on roller skates?
 3. Can you drive a car?
 4. Can you drive a boat?
- etc.

In each instance he grades himself as to the matter in question, assigning himself a value of 3 if he can do it very well, 2 fairly well, 1 if he knows something about it, and 0 if he knows nothing about it. The preliminary set of questions is followed by a more crucial set which the subject answers in the same fashion:

1. Do you know the letters of the alphabet in their order?
2. Do you know how to write decimals?
3. Do you know what a flywheel is for on a steam engine?
4. Do you know how a camera takes pictures?

Two or three weeks later, after the subject has presumably forgotten much of this preliminary test, he is given a test dealing with the information called for in the first case, such as:

1. What is the fourth letter after *M* in the alphabet?...
2. Write four-fifths as a decimal.
3. Flywheels are placed on steam engines in order to: aid in stopping them....help them keep going....tell how fast the engine is going.....

A considerable number of items of this type are used as a check on the subject's previous statements in an effort to determine whether he falsely overstated his ability in the first test.

Tests like this have the obvious limitation with adults that the subject may suspect the real nature of the test and not manifest his natural tendencies. Other efforts to measure deception by means of bodily changes that accompany emotion will be mentioned presently.

Example 59. In the same project tests of self-control employed the following. On each child's desk is placed a small box with a simple combination lock. The locks are set at a designated point at the outset of the experiment. The children are told that the boxes contain candy and that later on they will be told how to open them and get the candy but that they are not to touch them until directed. They then go about other classroom work, the boxes remaining on the desks. Proctors going up and down the aisles occasionally note the readings on the locks to determine the time elapsing before the pupil tampers with the lock.

Similarly, the pupils are given a booklet containing an interesting story. When it reaches a climax at the bottom of one page, large printed directions say, "Now go on to the next page and do the arithmetic problems before you finish the story on the page after that." By using pages that are treated with paraffin and folded in a certain way it is possible to determine whether the subject does the arithmetic at the time designated or whether he first completes the story.

Example 60. Tests of persistence have employed a stylus maze, that is, a metal plate with recessed pathways [13]. The subject is unable to see the maze because it is screened from his eyes. He starts at a designated point and with a stylus traces through the pathway to the exit. After three easy mazes he is given one which is impossible, i.e., has no correct passage, and he encounters blind alleys everywhere. He is graded by the examiner qualitatively on the basis of the response, ranging from 1, a subject who is careless and anxious to stop; 2, an excuse hunter; up to 8, a tenacious obstinate type; and 9, an analytical type.

Example 61. In one measure of aggressiveness, the subject looks the examiner in the eye while doing mental arithmetic [5]. The principle of the test is that an aggressive subject will look

the experimenter pretty steadily in the eye. The examiner counts the shifts of fixation. Furthermore, the distraction of looking at the experimenter would not retard the aggressive subject very much in his arithmetic, so the difference of speed in doing problems while looking at the experimenter and not doing so should give an index of aggressiveness.

Example 62. Pencil and paper tests of emotion were mentioned previously. Efforts have been made to measure the same characteristics more objectively by means of the bodily responses which accompany emotion—for example, when one is startled. Around his chest the subject has a pneumograph—a large rubber tube supported by a light spiral spring inside and connected by a smaller tube to a metal bellows. As the subject breathes, the air pressure in the pneumograph and the rest of the system changes so that the bellows expands and contracts. One end of the bellows is fixed; at the other end a small lever is fastened so that it can be actuated by the motion of the bellows, which is amplified mechanically. This lever carries a small barographic pen which writes on a polygraph or paper tape moving at a constant speed. In his hand the subject may hold a wooden handle with a delicate metal bellows on the end of it, weighted so that if his hand trembles changes of pressure occur in the pneumatic system; these are likewise recorded by means of another metal bellows. Blood pressure or some function of it may be recorded continuously by putting the usual rubber sleeve such as a physician uses around the upper arm, and inflating it with a moderate pressure but not enough to stop the circulation. This rubber sleeve leads to a rather complicated pneumatic step-down device which reduces the pressure so that it can be recorded by the usual type of metal bellows on the polygraph. A curve is traced which goes up as the blood pressure goes up, and vice versa. With a number of these devices attached, the subject is presented with various emotional situations—for example, a revolver shot or an electrical flashover—to see how quickly the record comes back to normal. The time required for this may be diagnostic.

This same general type of procedure supplemented with a careful oral examination is the basis of some methods of scientific crime detection. If a person is asked questions and lies about his answers, the emotional aspect of lying may be sufficient to

affect the blood pressure and produce certain irregularities in the breathing. In the hands of an expert this technique is useful in preliminary examination of criminal suspects. It is of interest in the present connection, however, because some financial organizations have actually employed experts to examine prospective employees by these methods, questioning a man about his previous financial history and his conduct in other similar positions. Contrary to what might be expected, if he does admit some previous financial irregularity and the polygraph record indicates that he is telling the truth, he is hired because if they know he has made some mistake in the past and he knows that they know it, he is pretty certain to be careful in his financial transactions in the new position. On the other hand, a man who denies any past financial irregularity but whose polygraph record looks suspicious is less apt to receive the job.

Example 63. Mention should be made of a few techniques that are still experimental and have never been used in an industrial situation but that may have some promise. The first is the effort to measure recklessness. It centers around a test procedure in which the subject is ostensibly trying to do one thing but actually is scored on some other aspect of the results which would not occur to him as significant. In one test the subject has a steel rod 6 feet long and half an inch in diameter which he is to balance, with one hand, on a metal plate on the floor so that when he lets go, the rod will stay in the air as long as possible before it falls over. The upper end is in the center of a hoop a foot in diameter with which the rod makes contact as it falls. A circuit through the rod, the metal plate, the hoop, a thimble on the subject's finger, and some relays operates electric counters which have a periodic interrupter in the circuit. Time intervals such as the length of time the rod remains in the air can be recorded in tenths of a second. The subject thinks he is being scored on the time the rod stays in the air, but the examiner actually is interested in how long he holds on to the rod before he lets go. Some subjects grasp the rod, release it, grasp it again, and repeat this several times before they are finally satisfied and release it, whereas others make just one quick setting of it and stop immediately. The supposition is that the latter individuals are more reckless or careless.

A similar test involves filling a number of small graduates with water from a beaker which is mounted on the end of a long handle to make the task more difficult. The subject is ostensibly filling the graduates to a point which is marked on the side. However, the examiner is interested in whether he fills them with one "shot" or whether he puts in a few drops at a time as he approaches the critical point. Still another test in this group involves a stylus maze which the subject traces visibly. He has numerous choice points where he may take a short pathway which is narrower or a long one which is wider. Every contact with the edge of a pathway is recorded electrically and also sounds a buzzer for the subject's benefit. The important score is not, as the subject thinks, how long it takes him but rather which pathway he chooses—whether he takes a chance on a short narrow pathway or "plays safe" and takes the long way around.

Example 64. A preliminary effort to measure susceptibility to frustration may be cited. The subject is presented with two lights differing slightly in intensity and is required to press a right or left key according to which light is the more intense. As successive pairs are presented, the test becomes increasingly difficult. Each mistake is signaled by a warning buzz and an electric shock. The series of increasing difficulty is continued until the subject has missed three trials in succession. Then he is returned to the easy ones at the beginning of the series and goes through it once more. The crucial point is whether on the second attempt the subject makes mistakes earlier in the series. A person who is somewhat inclined to frustration will be apt to be discouraged by the first experience and on the second attempt will break down earlier in the series.

The problem of personality measurement is obviously complicated and difficult. The question has been raised as to whether there are a few basic factors which account for most personality characteristics. One investigation along this line which may be mentioned [6] used a questionnaire dealing with personality items and then made a factor analysis of the results with 600 subjects. The first two factors that resulted were identified as nervousness or jumpiness and general drive, that is, pressure toward action. Two other factors could not be identified by consideration

of the factor loadings. A fifth factor was less clear but appeared to be a tendency to seek variety.

Objective personality tests are urgently needed by the personnel psychologist. As they become available they can be used to supplement capacity tests and increase the validity of test batteries.

SUMMARY

Mental tests may be classified according to whether they measure capacity or proficiency. The former deal with essentially innate factors, and the latter with acquisitions. The present chapter is concerned for the most part with illustrating tests of capacity of the sort that constitute the personnel psychologist's stock in trade. They may be further subdivided into tests of special capacity such as attention or memory and general capacity or intelligence. The conventional terminology used in dealing with such tests is justified on the basis of practical convenience. The main consideration is the extent to which the test correlates with the occupational ability which it is desired to predict; its name is in the last analysis irrelevant. Brief examples are given of tests for motor control, sensory capacity, attention, learning, association, memory, reaction time, space perception, reasoning, decision, ingenuity, and ability to follow directions.

Notions as to the nature of intelligence vary, but apparently some capacity is measured by our so-called intelligence tests that gives a person a poorer or better chance for survival in the economic struggle and that makes it possible in certain situations to predict occupational efficiency. This general capacity may be of the abstract type that is ordinarily measured in most tests; it may be of the mechanical type or even of the social type. Illustrations are given of individual and group tests of the abstract and mechanical sort and group tests of social intelligence. The scores attained in intelligence tests are usually handled by converting them into percentile scores for the group under investigation, or into terms of intelligence quotients.

The technique of factor analysis makes it possible to compute the intercorrelations of a number of tests and determine how many factors are needed to account for the intercorrelations. By inspecting the factor loadings it is possible to speculate as to the

nature of the factors. When this procedure is followed with a group of intelligence tests, seven factors are discovered and identified. These "primary abilities" may facilitate the construction of subsequent intelligence tests.

Vocational predictions made on the basis of capacity tests sometimes break down because of personality factors. The measurement of these latter is difficult. Personality tests in which the subject evaluates himself may be satisfactory in clinics, but in the employment office the individual will be tempted to answer in the way he thinks will insure him the job. Some such tests are described, however, together with some preliminary efforts to devise objective tests of personality.

REFERENCES

1. Allport, G. W. A Test for Ascendancy-Submission. *Journal of Abnormal and Social Psychology*, 1928, 23, 118-136.
2. Bernreuter, R. G. *The Personality Inventory*. Stanford University Press, 1931.
3. Edwards, M. E. *Social Intelligence*. M. A. Thesis, Ohio State University, 1931.
4. Garrett, H. E., and Schneek, M. R. *Psychological Tests, Methods, and Results*. New York, Harper, 1933, 235 pp.
5. Gilliland, A. R. A Revision and Some Results with the Moore-Gilliland Aggressiveness Test. *Journal of Applied Psychology*, 1926, 10, 143-150.
6. Guilford, J. P., and Guilford, R. B. Personality Factors N and G D. *Journal of Abnormal Psychology*, 1939, 34, 239-248.
7. Hartshorne, H., May, M. A., and Maller, J. B. *Studies in Service and Self-Control*. New York, Macmillan, 1929, 559 pp.
8. Hildreth, G. H. *A Bibliography of Mental Tests and Rating Scales*. New York, The Psychological Corporation, 1933, 242 pp.
9. Humm, D. G., and Wadsworth, G. W. The Humm-Wadsworth Temperament Scale. *American Journal of Psychiatry*, 1935, 92, 163-200.
10. Hunt, T. What Social Intelligence Is and Where to Find It. *Industrial Psychology*, 1927, 2, 605-612.
11. Kelley, T. L., Miles, C. C., and Terman, L. M. Ability to Influence One's Score on a Typical Pencil and Paper Test of Personality. *Character and Personality*, 1936, 4, 206-215.
12. May, M., and Hartshorne, H. *Studies in Deceit*, New York, Macmillan, 1928, 720 pp.

13. Morgan, J. J. B., and Hull, C. L. The Measurement of Persistence. *Journal of Applied Psychology*, 1926, 10, 180-187.
14. Moss, F. A., and Hunt, T. Ability to Get Along with Others. *Industrial Psychology*, 1926, 1, 170-178.
15. Paterson, D. G., et al. *Minnesota Mechanical Ability Tests*. Minneapolis, University of Minnesota Press, 1930, 586 pp.
16. Terman, L. M., and Merrill, M. A. *Measuring Intelligence*. Boston, Houghton Mifflin, 1937, 461 pp.
17. Thorndike, R. L., and Stein, F. An Evaluation of Attempts to Measure Social Intelligence. *Psychological Bulletin*, 1937, 34, 275-285.
18. Thurstone, L. L. *The Vectors of Mind*. Chicago, University of Chicago Press, 1935, 266 pp.
19. Thurstone, L. L. A New Conception of Intelligence. *Educational Record*, 1936, 17, 441-450.
20. Thurstone, L. L. *Primary Mental Abilities*. Psychometric Monograph. Chicago, University of Chicago Press, 1938, 121 pp.
21. Woodrow, H. The Common Factors in 52 Mental Tests. *Psychometrika*, 1939, 4, 99-107.

Chapter V

MENTAL TEST TECHNIQUE

The preceding chapter has given a notion of the types of mental tests that are available for a psychologist who is undertaking employment research. As previously mentioned, he needs to know the tools that are available and the proper ones to use on various occasions. But he requires in addition a skill in using the tools and a knowledge of many technical points that must be observed in test administration. A perfectly good plane in the hands of a novice will not produce a smooth plank, and a reliable and well-standardized mental test may yield worthless results if not properly administered. The present chapter will be devoted to test technique, with special emphasis on the methods of administration, the devising of test material, and the scoring of results. Most of the principles brought out will be applicable to tests in general, but where this is not the case they will be discussed from the point of view of personnel psychology.

METHOD OF ADMINISTRATION: INDIVIDUAL VS. GROUP TESTS

There are two methods of giving tests—the individual method and the group method. As their names imply, in the first, one person at a time is tested, while in the second a number of people take the test simultaneously. The individual method involves one examiner for each subject who is being tested at a given time. In the group method the number of people tested by one examiner is limited only by the number of seats and the acoustics of the place in which the tests are administered. The testing of 500 persons simultaneously is common.

Comparative Advantages. Each of these methods has its advantages and disadvantages. In the individual test the examiner is in a position to observe everything the subject does and if

anything goes wrong he is able immediately to make the proper adjustment. In a group of people being tested there are some who, in spite of all precautions to make the directions fool-proof and to administer the tests in standard form, get a bad start or do what they are not supposed to do. Such a simple thing as turning to page 6 after a specific order from the examiner to "turn to page 4" is frequent in a group. Some subjects will work at such a high level of attention that they will fail to see the word "stop" printed in bold-face type. If the examiner asks, "Does everyone understand what he is supposed to do?" some members of the group who do not understand will maintain respectful silence. But if the examination is given individually, the examiner will notice if the subject turns to page 6 instead of 4 and will correct the mistake instantly; or if he runs by the word "stop" he will immediately call his attention to the fact. If the subject does not understand the directions he will be more inclined to admit it when not in the presence of other subjects; at any rate, in his initial attack upon the test he will manifest his lack of understanding. The individual test, then, has a greater certainty that the subject will do what he is told, that he will get a proper start, and hence that the results will be typical of his ability under the prescribed conditions.

A second advantage of the individual test is that it provides more of a "clinical picture" of the subject. In a group test the examiner obtains no data except from scoring the test blank. There are occasions, however, when it is important to observe how the person goes at the test. If he attacks it with zest and apparent effort, his results are perhaps of some value, while if he goes at it listlessly and with apparent lack of interest, this attitude doubtless vitiates the test score but may be symptomatic of other things with which the examiner is concerned. A psychopathic subject under the pressure of the test situation may manifest emotional disturbances which he would not show under ordinary circumstances. If a certain portion of the test is not marked at all, it is impossible to tell, in the group method, whether the subject overlooked it, misunderstood, was unable to do it, lost interest, became frightened or angry, or had his attention distracted by a bird outside the window. While this "clinical picture" is usually more important in examinations given

to cases of suspected mental disease or mental defect, it is sometimes important in the employment situation. The writer was examining a man who had supposedly recovered from shell shock, with reference to employment on a fatiguing job requiring considerable patience and involving rather complex machinery. The man reacted normally at the outset, but in the course of the first test "blew up," protested violently against the tests, and manifested other psychopathic symptoms. Obviously, it would have been dangerous for him to undertake the work in question and he was given an unskilled job with simple implements outdoors. In a group test it is doubtful just what he would have done and it is certain that the results on his test blank would not have been as illuminating as his remarks. The person's extraneous reactions during the test are thus, in some instances, of interest and of practical importance.

A trained clinical examiner administering an individual test will be alert for any unusual behaviors such as those mentioned. To facilitate their more systematic observation he may even be provided with a check list to use in noting and interpreting behavior during the test. A list suggested by Bingham and adapted from Baumgarten is rather detailed [2, 229]. The list is divided into five portions; a few excerpts follow.

1. During the preliminary instruction: looking at the examiner; gazing around the room; asking questions; general approach such as active or interested; anticipatory remarks; and judgment or criticism of the task such as finding fault.
2. During the execution of the task: degree of apparent concentration; expressions of emotion such as surprise or displeasure; bodily movement, if the test involves some type of coordination; manner of work such as systematic, spasmodic; behavior as difficulties emerge—for instance, asking for help or immediately giving up; and conduct while being helped, such as indifferent or pleased.
3. Attitude toward his performance: whether he notes his mistakes and makes an effort to check the results.
4. Conduct at the end of the test: remaining silent and watching quietly or asking questions and showing interest in his performance.
5. After the test: leaving his materials, if any, in order or disorder; leaving the testing place quickly or slowly.

Such a detailed list may be used effectively by a trained ex-

aminer, but even he must be careful not to confuse what he observes and his own interpretation. The individual test in the right hands thus makes it possible to secure a lot of supplementary information about the person, in addition to the actual test score.

A third advantage of the individual method is that it permits greater flexibility in the selection of the tests. Some tests necessitate material equipment ranging from a picture puzzle up to an electrical device worth hundreds of dollars. In a group test every person must have the same kind of blank or apparatus, and if the latter is expensive it is often unwise to provide many duplicates, especially in the early experimental stages of the project. The natural result is a limitation in the tests that are to be tried out if the group method is used. In some problems, such as selecting clerical workers, this does not seem to be a serious drawback; but in analyzing some types of vocational ability, such as flying an airplane, it is highly desirable to evaluate rather complicated mechanical techniques. In general, the more tests tried the better final selection of tests for an occupation it is possible to make. The individual method affords this greater flexibility in selection.

Against this array of advantages of the individual test there is only one outstanding advantage of the group test—its economy of time. This is a tremendous advantage, however, in the practical situation. At Ohio State University every fall a two-hour test is given to about 3000 freshmen in one day by a small corps of examiners. If the test were administered individually, an examiner working on a reasonable schedule could finish the job in two or three years. In 1917-18 something like 100 examiners tested some 1,726,000 recruits within about a year. It would have taken one man between 600 and 700 years to do this job individually. In the practical situation it is necessary to set the aforementioned advantages of the individual test over against the saving of time and expense in the group test.

There is a scheme that is often used, however, to maintain some of the time-saving of the group test without sacrificing appreciably the advantages of the individual test. This involves the use of a small group—perhaps ten or a dozen. A group of this size may be seated at tables with space between them or in some other fashion so that the examiner by walking around the room

can look over everyone's shoulder. He can thus give almost as much supervision and make almost as careful individual observations as he would in the individual test. After he gives the signal to begin work, he can walk around rapidly; a glance at each paper will tell him whether everyone has started correctly and has apparently understood the directions. He can also note whether the subjects turn to the correct page or stop at the proper place, and observe numerous other things just as he would in the individual procedure. If anything is wrong he can almost immediately make the proper adjustments, such as assisting in finding the place or giving supplementary explanation where warranted and if necessary allowing extra time to compensate. The examiner can notice, moreover, many individual aberrations in attitude, because with the small group he can give a certain degree of attention to all of the subjects. He will doubtless "spot" anyone who is reacting in an unusual way and observe him more closely. In short, the first two advantages of the individual test may be obtained rather satisfactorily in the group test provided the group is small.

The other advantage of the individual test mentioned above, namely, the possibility of using more equipment, cannot be obtained in the group without considerable outlay. To be sure, duplicate sets of equipment may be provided and several subjects perform simultaneously. Mechanical assembly tests (cf. Example 47, *supra*) are sometimes administered to a dozen subjects in this manner. But this practice grows increasingly impractical as the equipment becomes complicated and expensive. However, a combination of the two methods is sometimes possible. Suppose that the entire program for each individual involves ten tests that employ printed blanks and two that require technical equipment. It can sometimes be arranged to give the tests involving blanks to the persons simultaneously and then have these subjects return individually for the two tests requiring apparatus. In testing applicants for a job, it is often possible to give them a portion of the test in a group and then let them wait while each is given his individual tests. In examining employees where it causes too much confusion to have each one leave his work twice to be tested, a certain amount of time can be saved by scheduling appointments so that two persons will always be taking the group part of the test simultaneously. For instance, the first man comes

and takes his individual test. Just as he finishes according to schedule, the second man enters and they take the group part of the examination together. The second man stays for his individual test. A similar procedure is repeated with the third and fourth men.

Comparative Difficulty of Technique. Mention should be made of a further difference in the individual and group methods from the standpoint of technique. The former usually necessitates a somewhat more skilled or better-trained examiner. The group test is usually somewhat more foolproof and somewhat safer in the hands of the untrained. This difference is not theoretically intrinsic to the methods. But in the tests that have been devised primarily for individual use, the examiner has to employ considerable tact and judgment in the course of the examination. In giving directions orally much depends upon the emphasis. One examiner might say, "*Work as fast as you can* without mistakes," while another says, "*Work as fast as you can without mistakes.*" The results would be altogether different in the two cases because the subjects were given an entirely different "set." In reading numbers to be memorized, the examiner has to control the time carefully. There is also the danger that he will put test instructions in his own words, thereby invalidating the standardization of the test. For example, one of the intelligence tests shows a picture of a broken circle. The subject is told that this represents a circular field with a gate and that a ball is lost in this field; he is to take a pencil and trace the path he would follow in hunting for the ball, assuming he entered through the gate and didn't know where the ball was. The purpose of the test is to see whether he has a systematic plan such as a zigzag back and forth across the field or a spiral arrangement. An unskilled examiner may unwittingly instruct the subject to take the pencil and "show how you would go around in the field to find the ball." Using this word "around" gives the whole thing away and actually suggests a circular or spiral pattern to the subject. The test is standardized on the basis of the assumption that the subject is not to receive this cue. Thus, the individual test does require greater care in its administration and is less foolproof. The ordinary group tests, especially those intelligence tests which have been published, are almost self-administering. About all the

examiner has to do is operate a stop watch and say "Begin" and "Stop" at the proper moment. All directions are printed on the blank so that the personal equation of the examiner does not enter. The greater necessity of having a skilled examiner does not apply to tests actually devised for group procedure but given individually.

In embarking on a testing program then, the decision as to what tests to use will depend somewhat on the ultimate organization with respect to the conduct of examinations. If the methods are to be left ultimately in the hands of persons without psychological training (a condition by no means desirable), it is unwise to introduce any individual tests of the sort that require a particular technique on the part of the examiner. In such a case it is better to adopt group tests or at least tests arranged in as fool-proof form as the usual group test.

Organization for Administration of Tests. Group and individual tests require a somewhat different organization for their administration. For the former, a room is needed that is large enough to seat comfortably as many as are to be tested. It is desirable to have sufficient space between the subjects so that they will not copy from one another's papers, or else provide the test in two forms of equal difficulty and distribute alternate forms to the subjects in alternate seats. In testing a large group it is further necessary to have assistants to aid in the prompt distribution and collection of blanks in order to insure that the subjects do not begin work before they are told or continue after the signal to stop. In the individual test, on the other hand, seating facilities are needed for only one subject, but space is required for whatever technical equipment is used. A room for individual testing often resembles a small laboratory. Usually the examiner can handle the individual test alone, although in some instances an assistant is desirable to take readings on the apparatus or to make notes of the subject's responses.

ADMINISTRATION: METHOD OF TEST RESPONSE

Oral Method. The subject may be required to make his response by various methods—oral, written, or performance. As their names imply, the subject may speak his answer, write it on the paper, or manipulate the test material in some other way.

In the earlier types of test the oral method frequently was necessary in order to time the response more accurately and obviate any error due to differences in speed of writing. This necessity has disappeared with the evolution of test techniques and the oral has been largely supplanted by the written method (*infra*). The oral method still is used, however, in giving individual tests such as Binet where some emphasis is placed upon the clinical aspects described above. It is also employed where the subject's literacy may be a handicap, as in some trade test projects.

Written Method. The written response is obviously necessary for group administration. To have a number of people give oral responses simultaneously would be absurd because of the difficulty of recording their separate responses—to say nothing of the noise. Many tests that originally required oral administration can be handled satisfactorily by multiple choice procedure in which, instead of giving his own free response to a word, the subject selects one of several alternatives. For instance, instead of giving the opposite of a word like “good,” he has an item like the following:

GOOD is the opposite of: NICE, FINE, BAD, POOR.

Here the time required to underline an alternative word is negligible and the test is essentially one of speed of association rather than motor performance. Where the nature of a test lends itself to this kind of arrangement the advantage of the oral over the written method of response disappears. There are also test situations in which, though actual words are written, the speed of writing does not introduce a serious error because the time spent in writing is slight compared with the time spent in deciding what answer to write. For instance, in a test comprising items like this:

A	E	U	E	U	O	A	E
	U	A	U	O			
		E	E	U	U	O	A A E U

in which the problem is to discover the relation of the letter O to the rest of the line that is the same in all three lines, the time spent in writing down the answer “after the second U” is slight compared with the time taken to discover this relation. In such

cases the written form of response is as satisfactory as the oral. Inasmuch as the written method is necessary in group tests and these group tests are desirable because of their time-saving, it is fortunate that these modifications in the technique of written responses have taken place.

Performance Method. In certain kinds of tests it is impossible to use either the written or the oral type of response. For instance, in assembling a picture puzzle the subject cannot tell verbally how to do it nor can he write out the method in detail. It is necessary for him to *do* it. Similarly, in assembling simple mechanical contrivances in tests for ingenuity, or in performing a series of complex motions in imitation of the examiner, it is necessary to have the subject actually make the motions. In measuring his reaction time he must press or release a telegraph key when he perceives a signal. There have been recent efforts to adapt some tests of this sort to the written form so that they can be given by the group method. Tests of the puzzle type sometimes have pictures of the loose parts numbered and the subject puts the numbers in the proper place on the blank to show where the parts belong. If the examiner touches a series of four points repeatedly in a complex order, the subject, instead of imitating him directly, may write the numbers of the points in the order in which they were touched. However, there will probably always be some kinds of tests which it will be impossible to adapt to a written form, and here the performance type of response will have to be maintained.

ADMINISTRATION: TYPES OF TEST RESPONSE

The type of response made by a subject depends on the way the test is organized. There are several possibilities. In the first place, it may depend on the wording of the question or item. For instance, the subject is given a list of words and required to give the opposite of each, or he answers questions of the sort: "Arm is to elbow as leg is to what?"

In the second place, the response may depend on the location of the answer. This is typified by the "completion" test in which words are omitted from a text and the subject supplies the missing words, as in the following: "In winter the**** is on the ground and the * * * * blows it into big * * * * *." Or the words

may be given with certain letters missing, like: "c*vw," "ho**e," "cl*ck," and the subject supplies the missing letters. The subject may or may not be informed regarding the number of letters omitted. The essential point is that his response is determined by the context and by the location of the answer.

The third type of response requires the subject to select his answer. He is provided with alternative answers from which he chooses the correct one. The number of alternatives may be 2 or more. The following illustrations are typical:

Good—bad.....same—opposite

A Zulu has Two; FOUR; SIX; legs.

Oyster : shell :: banana : TREE; PEEL; SIDEWALK; FRUIT.

An important consideration with reference to the number of alternatives is the possibility of getting the correct answers by guessing. With the two alternatives a person who knows absolutely nothing about the items involved and merely guesses at each will get approximately half of the items correct, just as in throwing a coin a large number of times approximately half of the throws will be heads. Hence, unless some allowance is made, an individual may attain a respectable score in such a test and apparently possess ability of the kind involved when this is not the case at all. With the three alternatives the chance of guessing the correct one is somewhat smaller—approximately one in three. In such an instance the score attained is more apt to represent the subject's actual capacity, although even here there is some possibility that accident will play into his hands. With four alternatives the probability of making a high score in the test by accident is rather small, and with five or six alternatives it is so remote that it is usually disregarded altogether. Tests with four to six alternative answers for each item are widely used.

A fourth type of response requires the subject to match items in two groups. For example, a test was mentioned previously in which on one side of the page are pictures of several tools including a hammer and screwdriver, and on the other side pictures of several items such as a nail and a screw. The subject selects items in the second group to correspond to those in the first group—for instance, matching the nail with the hammer. Or again two lists of proverbs are furnished and the subject locates

one in the second list that teaches the same lesson as a designated one in the first.

ADMINISTRATION: LIMITATION UPON TEST RESPONSE

Time Limit vs. Work Limit. Some limitation must obviously be placed upon the subject's responses in taking a test. He cannot work for an indefinite length of time, nor can he have unlimited material with which to work. Consequently, it is necessary to set either a time limit or a work limit. In the former all the subjects work for a constant length of time—e.g., four minutes—and are graded in accordance with the amount they accomplish in that four minutes. In the work limit they all finish a constant amount of test material—e.g., selecting 40 opposites—and the number of minutes and seconds required to finish the task constitutes the differential score.

Time Limit Preferable in a Group Test. The time limit and work limit are equally adaptable to statistical treatment of the results. The time limit, however, is generally to be preferred for a group test. It is possible to have a number of subjects work simultaneously for the same length of time and then subsequently score their individual accomplishment. If the members of a group are required to complete the same amount of test material, it is difficult to obtain a record of the time required by each individual to finish the test. This is sometimes attempted by placing a fast clock where it is visible to all of the subjects and starting them together and then having each one, as soon as he finishes, look at the clock and note the exact time on his blank. This procedure, however, implies honesty on the part of the subject. In the usual employment situation where a job may be at stake, it is dangerous to trust a person in this way. Unless the test is given individually so that the examiner himself can measure the time consumed, the time limit is to be preferred to the work limit.

Work Limit Feasible with a Long Test. It is usually not feasible to have the subjects bring their papers to the examiner as they finish and let him record the time. Most projects involve a group of tests each of which requires only a very few minutes. These may be given in succession, but the time for each must be recorded separately. If a test requires only one or two minutes for its completion, it is obvious that the time taken in bringing

the papers to the desk will make an appreciable increment. Suppose two persons finish simultaneously, but one is in a front seat and the other in the back of the room; the former may get a score of one minute, and the latter, one minute and fifteen seconds. This difference of 25 per cent will be entirely misleading. This sort of procedure is justified only when the time taken in actually completing the test is so large relative to the time taken in bringing the blank forward and having it recorded that the latter is negligible. If the test itself takes half an hour the fraction of a minute involved in getting the time record will be insignificant. One type of test is designed specifically in the light of the foregoing facts—the “omnibus” test (cf. p. 142 *infra*). In this type the different kinds of test items alternate throughout rather than appear in separate groups and the only score desired is the total time for all the items. With this sort of test the above procedure is justifiable and it is possible to give the test to persons who drop in at irregular intervals by merely marking on the blank the time they begin and the time they return the paper. In this way it is unnecessary to wait for a quorum before beginning to administer the test.

A similar situation prevails in cases like the following battery of performance tests of the form board and assembly type. The subjects are given all the parts spread out in a standard fashion, and each separate unit when properly assembled fits into a compartment in a large box. Closing the large box signals that the subject has finished. The performance requires from 15 to 30 minutes, so that if several people take the test simultaneously it is simple for the examiner to record the starting and finishing time for each one.

Comparative Reliability of Time and Work Limit. The comparative accuracy of the time and work limit has been studied empirically [9]. The problem can be formulated in this way. If a group of people take a given test with a time limit, is their comparative standing about the same as it would be if they had used a work limit? The test involved was a speed of reading test (Example 14, *supra*) in which the subject goes through a considerable number of brief paragraphs in each of which there is one wrong word. The number of these wrong words located per unit time constitutes the test score. Two forms of this test, A and

TABLE 11. COMPARISON OF TIME AND WORK LIMIT IN
SPEED OF READING TESTS¹

I. Form A Time Limit vs. Form B Work Limit	$r = .87$
II. Form A Work Limit vs. Form B Time Limit	$r = .84$
III. Form A Work Limit vs. Form B Work Limit	$r = .86$
IV. Form A Time Limit vs. Form B Time Limit	$r = .84$

B, of similar difficulty were available. Each subject took both forms, but some groups of subjects worked with a time limit, others with a work limit. Forms A and B were correlated for each group. Over 1000 subjects participated. The four comparisons made are indicated in Table 11. When one group is given time limit and the other work limit, the two correlate quite highly. In fact, they correlate just as highly as do two forms of this test both given by one method. This experiment suggests therefore that if results with this test are typical, work limit and time limit methods yield essentially the same results and that the decision as to which one to use may be made on some other basis, such as utility.

Determination of Proper Limits. The amount of material selected for a work limit test depends on two things. On the one hand, enough material must be used to give a fair sample of the ability in question. Half a dozen items may not be typical, while 100 may be little better than 75. This depends on the type of test. On the other hand, the amount of material is somewhat determined by the approximate length of time that can be devoted to the test. It is usually undesirable to include so many items that subjects will require several hours to finish that particular group.

With the time limit method it is important to determine in advance exactly what limit will be most satisfactory for the material that is provided. The general principle is that the time limit shall be such that the best individual will nearly but not quite finish the entire test. If many of the subjects finish the test, it is impossible to differentiate between their ability, for one may have barely finished, while another may have had a minute to spare and could have done a considerable number of additional items had they been available. On the other hand, if the best

¹ After Paterson and Tinker.

person finishes only half the items there is no need to have the other items on the blank.

Test construction usually necessitates preliminary experiments with a few people individually to determine the actual time required for each single item. It is common practice to select items on which the individual subjects agree fairly closely. With this information as to the difficulty of each item it is simple to arrange work or time limits. One can tell, for example, how long it will take on the average to complete 50 items, or how many items will be required to occupy the average subject for 10 minutes.

ADMINISTRATION: GENERAL PRECAUTIONS

Standard Conditions. A few precautions of a general nature are to be observed in giving tests. The examiner has to adapt himself to the conditions available with reference to many details such as the arrangement of materials and equipment, reception of persons to be tested, etc. One fundamental point, however, must be observed. All the subjects must take the tests under standard conditions. A chemical reaction does not depend appreciably on ventilation, room temperature, time of day, external noises, or nervousness of the elements involved. In a psychological laboratory or test room it is altogether different. If some subjects take tests when surroundings are quiet and others take the same tests when a freight train is being made up outside the window, the latter are at a disadvantage and the results are not comparable. The same is true if one group takes them in the morning when fresh. Likewise if one test room is well lighted and another has illumination of insufficient intensity or a distracting glare, results under the two conditions cannot be compared. If some subjects use pencils that are too hard and sharp and stick into the paper causing delay, a source of error is introduced. Psychological experiments reveal the extent to which rather slight changes in environmental conditions influence mental efficiency. Some individuals may be able to abstract from or ignore such things, but one cannot be sure he is testing such a person; the natural tendency is to be affected by distractions. Inasmuch as in a mental test the attempt is to measure one thing at a time, it is desirable to exclude other variables that may influence the re-

sults. Consequently, it is of importance to keep the test conditions standard and constant as far as possible.

Proper Attitude. Another general precaution that it is well to observe deals with the attitude of the examiner and the subjects. It is quite possible for the former to inspire an antagonistic or an alarmed attitude on the part of the latter. A subject who is resentful probably will not do his best and one who is frightened is liable to be somewhat distracted by the emotion. Consequently the examiner should at the outset establish rapport. This term was used originally in hypnotic technique, but has been aptly applied to mental test procedure. If A is hypnotized by B, he will accept suggestions from B and carry them out, whereas if C tells him to do something the suggestion will be less effective. This is explained by the fact that A and B are *en rapport* and A is more inclined to cooperate with B than with C. Similarly in giving mental tests the examiner should get the subject into this attitude of cooperation or, in everyday parlance, get the subject "with him." Under these conditions the subject will do what he is told, will do his best, and will try to conform to the wishes of the examiner. More extensive procedures along this line are possible with individual than with group tests. The establishment of rapport calls for tact on the part of the examiner, sometimes an explanation of the purpose of the test project (depending on the intelligence of the subjects), and a general atmosphere of cordiality. It is often well to precede the tests with a few moments of general conversation or with remarks leading up to the matter in hand, gaining the confidence and good will of the subjects and allaying suspicions or fears. Often a "shock absorber" is used for the last of these contingencies. This is a brief test which precedes the others and is not necessarily scored, but merely serves to get the subjects accustomed to the test situation. The examiner must adapt himself to circumstances; but whatever they may be, he should strive for rapport with the subjects and have the whole atmosphere of the examination one of willing cooperation.

Another procedure that often contributes to rapport is to give the subject an indication of how successful he is in some of the tests. The subject frequently raises that question himself: "How am I doing?" The examiner may be noncommittal, but if rapport is difficult to secure, some indication of progress may help. If a

shock-absorber test is given it is always feasible to tell the subject something about his score in that particular portion. Some examiners recommend the inclusion of irrelevant tests here and there in the series so that the subject can be told his scores in these with impunity. There is, of course, the danger of discouraging a person by indicating that his score is poor, or leading him to put forth less effort because he thinks he is doing unusually well. If his scores are at either extreme it may be inadvisable to inform him of it. If he persists in asking about his score the examiner may have to be a little indefinite and say that it is "pretty good" or something to that effect. In most cases judicious use of information about the subject's own performance may be helpful in providing better rapport and also in motivating him. (Cf. p. 75 *supra*.)

There is another point to be observed particularly by the inexperienced examiner. He should himself be thoroughly familiar with the test procedure before administering the tests. If he makes mistakes or has to change his directions after giving them, it is embarrassing, the subjects lose confidence, and it is liable actually to vitiate the results. He should rehearse his part in advance if necessary.

TEST MATERIAL

Difficulty of Material. In devising material to be used in a particular mental test, one thing that must be considered is the difficulty of the test items. Most tests comprise a considerable number of separate items of the same general sort, e.g., 30 examples of opposites. These should not be made up and used at random, but rather the difficulty of each separate item should be determined. This is usually done, as suggested above in connection with setting time or work limits, by experimenting individually with a number of subjects and measuring the time taken to do each single item. If the results for the various subjects show fair agreement with one another, the average time for an item may be taken as an index of the difficulty of that item.

Speed vs. Power Tests. Assuming that the difficulty of the various items is known, there are two different trends in test construction—to arrange the test so that all of the items will be of approximately equal difficulty (speed test) or to have them

increasing rather uniformly in difficulty (power test). In the first of these the interest is in the amount of performance per unit time, while in the second it is in the ultimate difficulty of performance that can be attained.

The speed test may be typified by a page of random numbers in which all pairs of adjacent numbers whose sum is 10 are to be canceled. All such pairs will be of approximately equal difficulty and the number of pairs canceled in 5 minutes may constitute the individual score. Consequently, if one person scores 100 and another 125, it may be stated that the latter is 25 per cent superior in this sort of performance. Almost any kind of test may be given in this speed form, provided enough items of equal difficulty can be devised. It is most frequently used in situations where the interest lies in the subject's alertness or ability to think or act quickly. It intentionally and avowedly puts a premium on speed.

The power test may be typified by a number completion test in which a series of numbers are given and the subject is required to complete the series. The items may start with relatively easy ones like:

1 2 3 4 5 6 7

and lead up through gradually increasing degrees of difficulty to items such as:

2 4 8 3 9 27 4

Power tests are usually given with a time or work limit, but the temporal aspect is not regarded with as much concern as in the speed test. If a time limit is set it is usually such that the subject will get about as far along in the items of increasing difficulty as he would if he had unlimited time. While he might do a few more items if he had an opportunity to take the blank home overnight (subjects occasionally make this request), he would not do very many more, and the number of items he passes under the test conditions is a pretty fair indication of his proficiency in this particular sort of task. The power test is most often used in situations in which interest is not in a person's intellectual alacrity, but rather in his ultimate possibilities of intellectual attainment.

There is a popular misconception that should be cleared up in this connection, namely, that the speed test is not a "fair" test. The subject states that if more time had been allowed he would have been able to do better, and that he has known persons who would be very slow in thinking out items of this sort, but who were nevertheless economically and socially successful. Of course the subject might do more with unlimited time—and so would his competitors. But the purpose of the speed test is to find out not how much he can do at leisure, but how much he can do per unit time. As mentioned previously, the tests are so constructed that few persons will finish, in order that the scores may scatter over a considerable range. To be sure, much of our work in daily life is not done to the time of a stop watch, but it is true in general that the brighter minds work not only better but more rapidly. After all, the "fairness" of a test depends on whether it may validly be used in predicting some correlated capacity. If scores in power tests are more closely correlated with proficiency in clerical work than are scores in speed tests, the former will be "fairer" to use in selecting clerical workers, and vice versa. As a matter of fact, statistics show that the abolition of time limits would in many cases be disastrous, for there are definite tendencies for those who are proficient in tests which emphasize speed to make better messenger boys, clerical workers, engineers, and to rise in general to occupations on the business or professional level rather than on the level of unskilled or semi-skilled labor. Where tests devised for a practical purpose, such as predicting engineering aptitude, were given with and without time limits, their diagnostic value was greater in the former case [14, 275].

An exception to the foregoing is the experience with an intelligence test at a large university. It is designed to predict university scholarship and is revised from time to time in order to increase the validity of that prediction. It is a power test, for the items increase in difficulty. Some students were given two forms of the test, one with a time limit and the other with a liberal work limit [16]. The results are given in Table 12. The later forms of the test have the larger numbers. The coefficients represent the correlation between academic marks and intelligence test score. The data in a given row are for the same subjects. With one

TABLE 12. VALIDITY OF TIME LIMIT AND WORK LIMIT INTELLIGENCE TESTS²

Time Limit		Work Limit		Number of Subjects
Test Form	<i>r</i>	Test Form	<i>r</i>	
12	.38	9	.54	110
12	.46	10	.43	115
12	.44	11	.52	122
9, 10, 11	.32	12	.40	83
14	.57	15	.65	71
14	.43	16	.65	57
Average	.46		.53	

exception, the correlations are a trifle larger for the work limit administration. Inasmuch as the validity of the test has increased with each revision, the present results must be scrutinized to be sure that the larger correlations for work limit are not due merely to the use of a later, i.e., more valid, form. This is not the case, however, because in three comparisons the work limit involves an earlier form of the test and in three a later form. When time limit was correlated with work limit for each of the six groups, the coefficients ranged from .59 to .82 and averaged .74. As a result of this investigation the work limit method was adopted as standard procedure for this test. Those in charge state that the attitude of the students taking the test is better under these circumstances.

Selection of Misleads. In items of the alternative answer type it is desirable to give some attention to the incorrect alternatives or misleads. An item like the following would be absurd: "Columbus is the capital of: Ohio, Napoleon, Christmas, Lake Ontario." The three incorrect alternatives are so inane that even a person who is uncertain about his state geography can easily guess the correct alternative. It is possible to study the misleads empirically in order to determine how misleading they actually are. The procedure may be illustrated in the develop-

² After Workman.

ment of a vocabulary test [8]. This test involves a word and five alternatives, one of which is synonymous with it; for example:

He is CROOKED in business—FAILURE, COWARDLY, SUCCESSFUL, TIMID, DISHONEST.

This item was given to a considerable number of persons and the frequency noted with which each of these alternatives was marked. Four per cent of the subjects marked "failure," 1 per cent each marked "cowardly" and "successful," nobody marked "timid," and 94 per cent marked "dishonest." "Timid" obviously misled nobody and might just as well have been omitted altogether. Consequently it was changed and the item submitted to another group of subjects. The misleads were changed one at a time until they were all of about equal difficulty and not too similar to the correct answer in this respect. To cite another item:

A ROBUST child—ROLICKING, FAT, STURDY, CUNNING, NAUGHTY.

The frequencies of these five alternatives in order were 13, 25, 47, 2, 12 per cent. "Fat" misleads 25 per cent of the people; only 47 per cent of them get the correct answer, "sturdy." Even "rollicking" and "naughty" mislead rather large percentages. Changes were made so that the misleads were about uniformly deceptive.

Another possibility is to standardize the items with persons in the upper range of abilities of the particular kind under investigation. A good alternative should presumably mislead the average subject more than it misleads those at the upper level. In the preceding example the frequencies were 2 per cent, 16 per cent, 81 per cent, 0, 0, with a group of subjects in the upper one-fourth in vocabulary.

Location of Alternatives. Another variable to consider in multiple choice tests is the actual location of the correct alternative in the group. It is possible that when people are uncertain about an answer they are more inclined to guess at alternatives in a certain position. This possibility has been investigated [1] by using a test in which the subject had practically no clue as to the correct answer and essentially guessed on the whole test. Position number one was chosen most frequently, five next, followed by three, four, and two, in this order. The implication of this finding is that the correct alternative should be in different posi-

tions in different items. If it is in one of the preferred positions a subject who does not know the answer stands an undue chance of guessing it.

Arrangement of Test Material. The usual procedure in assembling tests is, as implied above, to group together items of a given sort. In a battery of special capacity tests these groups are kept entirely separate. Even in an intelligence test which comprises several different kinds of items it may be desirable to evaluate them separately. This is obviously facilitated by grouping them so that, for instance, the "attention test" and the "memory test" are entirely separate. Each group of items is generally preceded by the directions or instructions for dealing with those items. Usually the blank is so arranged that one set of items occupies a page and with its directions forms more or less of a unit. In the Army Alpha intelligence test, for instance, the first page comprises 12 items for which verbal directions are given and is labeled Test 1; on the next page is Test 2, comprising 20 simple arithmetical problems with directions printed at the top of the page; the next page constitutes Test 3 on practical judgment, comprising 16 questions with three alternative answers to each; etc. Each page lists test items of a separate kind with directions at the top of the page. This is typical of most scales or groups of tests used in experimental work—separate administration of different kinds of test items.

Omnibus tests, however, depart from the foregoing arrangement—and are designed to facilitate test administration. Instead of all the items of a given sort being arranged so that they occur together in a single test, to be followed by all the items of another kind grouped together, each test with its time limit of a few minutes, the items of the different types are intermixed in one way or another, with a single time limit or even a work limit for the whole. A typical one starts with three arithmetic problems followed by three practical judgment items followed by three disarranged sentences and so on; it then returns to three more arithmetic items, three more practical judgment items, etc. The choice of three successive items of a given kind is arbitrary. It might have been only one or it might have been 10. In some instances the experimenter is interested in providing quick shifts of attention from one sort of thing to another in order incidentally to

measure this factor as well as the general ability manifested in the test itself. Sometimes he is more concerned with getting the subject well started on one type of item before the shift occurs in order to note his ability to change his "set" after it is well established. The items may even be given in a random order. In the omnibus test the directions or explanation for all the kinds of items involved must necessarily precede the test proper.

The omnibus test, like that with the items grouped, may use items either of equal difficulty or of increasing difficulty. When all the items of a given kind are approximately equal in difficulty, the test is called a *cycle omnibus* test. When the items of each sort increase in difficulty throughout the test—i.e., each item is more difficult than the preceding item of that type—the test is called a *spiral omnibus* test.

Alternative Material. In making up test material it is well to devise additional items at the outset in order to provide alternative test blanks. If the original blank is in use for a considerable time, it is quite possible that a copy will get outside so that some persons will have access to it before taking the tests. Moreover, those who are examined will remember some of the items and discuss them with friends. Everyone engaged in a test project of any magnitude feels occasionally that some of the people who come in to be tested are not as naïve as should be expected. A subject not infrequently registers pleasure on recognizing items that are familiar and on which he has been "primed." When this situation arises, it is desirable to have another test blank involving different items, but of the same difficulty as the first. Other persons can then take the tests without profiting by any information they may have received previously, yet their results will be directly comparable with those of the people who have been tested previously. It is common practice to provide more than twice as many items as are necessary for one form when devising and determining the difficulty of the test items. Then if the data representing difficulty are available for all items, it is comparatively simple to select two groups of items of the same total difficulty. These two forms of the test are likewise of value in cases where it is necessary to test large groups of individuals crowded together so that there is danger of their copying one another's papers. The blanks may be distributed in such a way that subjects in adjacent

seats have different forms. In fact, most of the larger test projects issue a given test or scale in two or more alternative forms to provide for the contingencies above indicated.

Sensitivity. The test material should be selected with a view to sensitivity. A sensitive test is one that gives a considerable range of test scores with the group studied or that reveals marked individual differences in performance. If everyone taking the test scores either 29 or 30 points, it is not considered a sensitive test, whereas if some individuals score as low as 10 points and some as high as 80, the test differentiates clearly between the various subjects. In securing this sensitivity there are two things to consider. In the first place, the test should have many items or increments. Suppose a test involves only three items; it will be possible for the subjects to score 0, 1, 2, or 3 points. The best that can be done is to divide the subjects into four degrees of ability. In the second place, the items should be selected so as to be differential with the group studied. It is possible to have the items all so easy that everyone can do them as rapidly as he can write or make the appropriate marks. If, for instance, a group of college students are given a test of the order of 2×3 or $3 + 5$, they can do the problems as rapidly as they can write the answers, and it is probable that they will all make approximately the same score and hence the test will not be sensitive. On the other hand, a group of persons of low intelligence may be given questions that are so difficult that none of them will be able to do more than one or two. If, however, the difficulty of the questions is neither too little nor too great for the individuals being examined and if there is a sufficient number of questions, the test will be sensitive and reveal the desired individual differences.

TEST INSTRUCTIONS

Standard Instructions. The instructions given the subjects are almost as important as the test material because they must insure that the subject will do what the examiner actually wants him to do. Perhaps the most important point about instructions is that they must be kept standard or constant. If one person is told to do one thing and another person told to do something else, obviously their test results are not comparable. If one blank says, "Work as fast as you can," and another says, "Make no mistakes,"

quite different attitudes will be evoked and an altogether different emphasis on speed or accuracy given to the subjects. The second blank may show greater accuracy than the first, not because the individual using it is naturally more accurate, but because he is told to be more accurate. If one subject is instructed to complete every item before passing to the next and if another is told to skip any items which he cannot solve in a few seconds, the first may spend half the test period on a single item which he finds difficult, while the other may make a much higher score simply because he selects the items which he can solve easily. The emphasis on factors like these must be determined by considering whether the examiner is more concerned with speed or with accuracy in the particular problem for which the test is to be used. But the essential point is that once the instructions have been determined upon, they must remain constant for everyone who takes the tests. Sometimes, of course, supplementary explanation is given if the subject does not understand the original instructions. It is well to have this standard likewise so that no subject will be given any unfair advantage because of some implication in the wording. As a matter of fact, ideal instructions will need no supplementing, at least with adults of normal intelligence.

Clarity of Instructions. Another requisite of instructions is clarity. If they are ambiguous or incomplete so that the subject does not understand exactly what is wanted, they fail of their purpose. It is not safe for the examiner to compose the instructions and use them at once. It is highly desirable to try them out on a few persons, preferably of the type with whom the tests are to be used. Instructions that seem absolutely fool-proof to the one who writes them will frequently have some point that can be misinterpreted or some contingency that is not covered. If the subject is told, for instance, to "mark the correct word in each line," he may underline it as the examiner intended or waste his time making elaborate rectangles about the words. He may work up and down the page, although it was assumed that he would do the obvious thing and work across. If he is told to "cancel the vowels" he may be in doubt as to whether *w* and *y* are respectable vowels. Any number of minor points of this sort will come out in using test instructions. Hence it is well to give them to a small

experimental group and note any questions that are asked and any uncalled-for performance that results. The instructions can then be modified accordingly before being put to practical use.

In insuring the clarity of the instructions it is necessary to consider the general mental level of the persons who are to read or hear them. The vocabulary for people of low intellectual status must necessarily be simpler than that for persons higher in the scale. Statistical studies reveal a greater incidence of polysyllables and long sentences in "high-brow" magazines. Persons of lower status are likewise apt to require more detailed explanation. For instance, a group of college students, if given a page of numbers in random order and told to "cross out every pair of adjacent numbers whose sum is 10," will probably be able to do it, whereas a group of unskilled laborers will become paralyzed or profane. It will be necessary to tell the latter: "Wherever you see two numbers side by side that would give 10 if you added them together, draw a line through those two numbers. Remember that they must add up to 10 and that they must be side by side with no other number between." A safe rule in devising instructions is to step them down to the level of the person with the lowest mental capacity who is apt to take the test. The others may be a trifle bored, but this will not vitiate their results. It is better to play safe and insure that even the poorest one in the group understands what he is to do.

This necessity of adjusting the instructions to the type of subjects involved sometimes manifests itself in an unusual fashion. A form board test was being administered to a group of highly skilled machinists. They were told to put the blocks into the holes "where they would fit." In the usual construction of such boards the blocks are a little smaller than the holes so that there is perhaps a $\frac{1}{16}$ -inch clearance all around in order to facilitate manipulation. However, these particular subjects were men who had been working with micrometers and fitting things to $\frac{1}{1000}$ of an inch. When they found a block that went into a hole but had a $\frac{1}{16}$ -inch clearance it did not "fit" in their sense of the term and so they tried to find a hole with less clearance. This delayed them unduly and their test scores were useless. When giving the test subsequently to such persons the term "fit" was clarified.

Form of Instructions. The actual form of the instructions natu-

rally varies with the test involved. However, most instructions embody three parts—explanation, illustration, and practice. Some test material is usually presented to the subject while explanation is made as to what is to be done. Then this material is marked by the examiner by way of illustration, or else these or additional examples already marked are presented for study. Finally, further unmarked items are given for practice before beginning the test proper. While the subject may think that he understands the test from looking at the illustrations, he may find it a different matter when he comes to work out practice items himself. If he accomplishes these latter, it is certain that he understands what he is to do in the actual test. The following excerpts from the directions preceding a group omnibus intelligence test illustrate these three stages of explanation, illustration, and practice.

Inside this booklet you will find a lot of things to do. Samples of the different things to be done are given below, along with a few examples on which you can practice. You will be given plenty of time to study the directions and do the practice examples. These do not count as part of the test but are merely to make sure that you learn to do each kind of problem correctly.

1. GOOD is the opposite of: EXCELLENT; CHEERFUL; BAD; WRONG; TRUE.
2. LITTLE is the same as: SMALL; COARSE; PRODIGIOUS; FEEBLE; IMMENSE.

Underline one of the last five words in each line that makes the best sentence. If more than one answer seems correct underline the one that is the most nearly the same or opposite according to specifications. Mark only one in each line like this:

1. GOOD is the opposite of: EXCELLENT; CHEERFUL; BAD; WRONG; TRUE.
2. LITTLE is the same as: SMALL; COARSE; PRODIGIOUS; FEEBLE; IMMENSE.

Do the following problems for practice:

1. THICK is the opposite of: HEAVY; LARGE; THIN; SMALL; NARROW
2. SHY is the same as: BOLD; COY; FRIGHTENED; TIMID; SHINY.
3. CARELESS is the opposite of: NEGLIGENT; UNEASY; ANXIOUS; UNCONCERNED; CAREFUL.

-
1. a eats wood cow grass.

2. birds swim feathers have all.

The words "a eats wood cow grass" in that order do not make a sentence but they would make a sentence if put in the right order, only there would be one word left over. The sentence would be "a cow eats grass" with the word "wood" left over. The thing to do is to cross out this extra word "wood," like this:

1. a eats ~~wood~~ cow grass.

The words "birds swim feathers have all" would make a sentence if put in the right order, "all birds have feathers" with the word "swim" left over. The thing to do is cross out "swim" like this:

2. birds ~~swim~~ feathers have all.

Do the following problems for practice:

1. dogs climb meat eat.
2. Florida in cotton button grows.
3. ocean house in live fish the.

-
1. 2 4 6 7 8 10
 2. 32 20 16 8 4 2

Each number is derived in a certain way from the numbers coming before it. Study out what this way is. You will find in each problem one extra number that does not belong there. Cross it out like this:

1. 2 4 6 ~~7~~ 8 10
2. 32 ~~20~~ 16 8 4 2

Do the following problems for practice:

1. 22 24 26 28 29 30
2. 13 12 11 10 9 7
3. 1 2 4 16 64 256

-
1. Sky : blue :: grass : TABLE; GREEN; WARM; BIG.
 2. Locomotive : train :: horse : BICYCLE; HUB; BUGGY; BAGGAGE.

The first word "sky" is related to the second word "blue" in the same way as the third word "grass" is related to one of the words following it. You are to underline the word that is related to the third word as the first two words are related to each other. In this example "sky" is related to "blue" as "grass" is related to "green" because the sky is colored blue and the grass is colored green. Therefore "green" should be underlined like this:

1. Sky : blue :: grass : TABLE; GREEN; WARM; BIG.

In the second example, "locomotive" is related to "train" as "horse" is to "buggy," for a locomotive pulls a train and a horse pulls a buggy. Therefore "buggy" should be underlined like this:

2. Locomotive : train :: horse : BICYCLE; HUB; BUGGY; BAGGAGE.

Do the following problems for practice:

1. Bird : Sings :: Dog : FIRE; BARKS; SNOW; FLAG.

2. Eat : bread :: drink : WATER; IRON; LEAD; STONE.

3. Arm : elbow :: leg : FOOT; KNEE; SHOE; PICK.

In some cases, of course, the test is so simple that elaborate instruction is unnecessary—for instance, "Solve the following arithmetical examples," or, "Cross out every letter A on the page." However, a bit of initial practice is always desirable.

Practice. It is well to consider at this point the general question of how much practice to give in a test, inasmuch as such practice is part of or immediately follows the test instructions. Some studies have been made of the effect of practice on various motor or mental tests. It varies, of course, with different tests. Viteles reports on some mechanical tests including mechanical assembly, Minnesota Paper Form Board, wiggly blocks, packing spools in a box, and several others [15]. He was interested in correlations between scores made earlier and later in practice; in other words, whether the persons who did well in the test at the outset likewise did well later on. For some of the tests, the correlations were none too high. In a type of discrimination test the correlation between an 18-minute and a 2-hour session was .69, and in packing spools a similar correlation was .53. When the tests were continued for several hours and the first half-hour was correlated with the second, the first with the third, the first with the fourth, and so on, the correlations decreased progressively. This study suggests, then, that the initial scores in the test do not indicate very accurately a person's final standing after considerable practice. This does not answer, however, the further question as to whether the initial or the final portions of the test are more valid. This point will be discussed in a moment.

Another study investigated a wider variety of tests [5]. These tests were repeated four times with college students. The general trend was as follows: Tests of speed of movement and threshold discrimination showed little improvement with practice. Memory span for digits, easy cube designs where a person had to estimate how many cubes there were in a picture of a pile of them, and accuracy of movement showed about a 10 per cent improvement.

Simple manipulation and perception as in the Minnesota Spatial Relations Test and easy pencil mazes showed 50 to 70 per cent improvement. In tests where a technique for solution could be developed by repetition such as a difficult pencil maze, or in a rational learning test in which the subject simply had to learn what is essentially a code, the improvement was 100 to 150 per cent.

These two studies, while not sufficiently extensive to warrant generalization, do indicate that practice is an important consideration in test administration. The personnel psychologist would certainly be taking a chance if he started on a test program without knowing something about what happens to the test with repetition. Obviously, what is necessary is to give this test to a small group of people repeatedly and watch the practice curve, i.e., determine at what point there ceases to be any further improvement. With this information available there are several possibilities. In the first place, if the practice effect is not too pronounced it may be feasible to pursue the test to the limits of practice before taking the record that is actually used in the employment program. This is satisfactory if it does not require too much time. In the second place, if the practice effect is very slight, it may be altogether disregarded. Finally, it may be possible to select arbitrarily some particular stage of practice at which always to take the record for employment purposes. For instance, one may know just how far people progress in ten practice trials and give that many arbitrarily to everyone. This procedure is defensible only if the test is one on which the subject would not be apt to secure any practice or coaching prior to the examination. The crucial point then would be the validity of the test at that particular level of practice.

Two other considerations indicate some merit for this last suggestion. Many studies have been directed to the problem of whether practice increases or decreases individual differences. Obviously a test is desired which will scatter the subjects over a rather wide range of scores so that differentiation can be made between them. A review of a large number of these studies indicates that on the whole subjects vary less after practice than before [11]. This was true in 58 out of the 70 studies. Another consideration is that it actually has been demonstrated in some

cases that unpracticed tests do have considerable validity. A British investigator reports that some functions are more accurately measured by tests "repeated no more than is necessary to secure full understanding of the examiner's requirements" [7]. At any rate, it is obvious that the practice effect in tests is a moot consideration and psychologists should not proceed to use a test without having made some investigation of the extent to which it is influenced by practice.

Printed vs. Oral Instructions. Instructions may be oral or printed. Where the subjects are working with test blanks, it is current practice to print the directions on the blank. This has the advantage of eliminating the personal equation of the examiner. The subjects examined at various times are given exactly the same wording with no difference in the oral emphasis. In some cases, of course, printed directions are undesirable. Some subjects cannot read but can take performance tests; obviously the instructions must be given orally. Sometimes a limitation on printing or mimeographing service makes it necessary to economize by omitting the printed directions. Sometimes the oral method is used to prevent the subject from working ahead in the blank prior to the signal, although this difficulty can usually be avoided by alert proctors or by arranging the blank so that when working on one page the adjacent page is upside down. If oral instructions are to be used, effort should be made to keep them as constant as the written ones. Most examiners have the instructions written and actually read them from their copy or memorize them and give them verbatim.

INCENTIVE

Maximum Incentive. Incentive is a factor that must be controlled. This may often be done through the test instructions and hence it is discussed in the present connection. If incentive is not controlled, it introduces another unnecessary variable, and this is contrary to scientific method. If a chemist is studying the relation between the pressure and the volume of a gas, he does not let the temperature vary at random but keeps it constant so as to determine the relation between the other two variables. Similarly, a psychologist studying the relation between intelligence and vocational aptitude tries to stick to those two variables and keep

other things constant. If two persons take the same test and one does the best he can and the other does not try, another variable is immediately introduced. Their scores may be altogether different, although they have, perhaps, the same actual ability. Incentive, therefore, should be a constant rather than a variable and the only practical way of keeping it constant is to keep it at a maximum. Under these latter conditions we can say that one subject makes a certain number of points when he is doing the best he can and that he is superior to another who is likewise doing his utmost.

Securing Incentive. It is often possible to obtain this incentive by emphasizing in the instructions the importance of doing well. The exact statements used in introductory explanation of the purpose of the tests will vary with the circumstances, but the final statement that "It is important for everyone to do his best" is usually quite effective. In testing applicants for a job, incentive, of course, will take care of itself, because they realize that their score may have something to do with their being hired. In testing employees for research purposes the problem is more difficult. It may be that there is a possibility of the tests being used for promotion or readjustment of some sort and that it is desirable to let the subjects know this fact. Sometimes there may be an appeal to their pride, to the effect that "We are standardizing these tests and we want to find out what people who are already on this job and making good can actually do in the tests." With more intelligent subjects it is sometimes wise to explain the actual research problem and to enlist their cooperation in a scientific experiment. Occasionally competition may be used as a motive, such as the statement that so-and-so "broke the record on this test, now see what you can do." In a small group, if after one test the subjects compare notes, such as "How far did you get?" and this can be permitted without danger that anyone will work overtime, this may serve as an additional motive for the following tests. Competition with oneself is also effective at times. If a test comprises several parts, the subject may be urged in the second part to see if he can beat his record in the first. In individual tests favorable comment on test results will often motivate the subsequent tests. In the discussion of rapport (*supra*, p. 137) we noted that some unimportant tests may be included in the program for

the express purpose of telling the subject his score in order to increase motivation on the other tests. The particular kind of incentive that will prove most effective will depend on the type of subjects, the test situation, and the nature of the test. The examiner must adapt himself to these and strive for some effective means of keeping incentive at a maximum.

SCORING OF TESTS

Unequivocal Scoring. In devising tests consideration should be given to the possibility of unequivocal and simple scoring of the results. The first of these is in the interest of reliability, and the second in the interest of time-saving. The unequivocal character is necessary in order to insure that when the tests are scored or administered by various individuals comparable results will be obtained. If it is necessary, for instance, to determine whether the answers to certain test items are good, average, or poor, different examiners will doubtless differ in their judgment. The personal element will enter and different persons scoring the same test blank will arrive at a different total. If, however, the items have each a single correct answer, all examiners will obtain exactly the same score for a given subject. Hence it is desirable, whenever possible, to have the items of the single-answer type, whether this answer is given orally, written on the blank, or selected from a list of alternative answers.

Some tests are still in use in which the scoring is subjective, but an effort is made to standardize it. If the subject copies a geometrical figure or writes something to exhibit his own handwriting, recourse is had to some form of rating scale. This consists of a series of specimens of the geometrical figure in question or of handwriting, ranging from very poor quality to very good. These specimens have been standardized and each is assigned an appropriate number of points. In grading the test, the scorer compares each item in question with the specimens in the scale, determining which of the latter the former most resembles and assigning it the corresponding number of points. With a little practice in the use of such scales fairly reliable results can be obtained. All of the arguments, however, are in favor of the entirely objective and unequivocal type of score wherever the test can be adapted to this form.

Ease of Scoring. The ease of scoring from the standpoint of clerical work is especially important in written tests. In the oral test the examiner usually notes the scores on the items during the progress of the examination, but in tests given to large groups the clerical work of subsequently scoring the blanks mounts up tremendously unless effort is made to simplify the process. One fact that makes for great simplification is that in the printed blank it is possible to have the answers in the same location on each blank. If the answers are to be written, dotted lines or brackets may be provided, thus insuring the location of the answer. If the response consists of crossing out or underlining something, its location is already determined. In the case of the written answers, if they can be arranged in a column down one margin, a card containing the correct answers likewise in a column may be aligned alongside and the two columns easily compared. If the answers are simple symbols such as x and o , it is often easier to memorize the sequence. This may be facilitated by arranging the items in such a way that the correct symbols occur in rhythmical sequence, such as "xooxooxoo." Even with answers which consist of a list of words or numbers or letters, it is often rather simple to memorize them; this frequently takes place incidentally after the key has been used for a time in correcting the blanks. In case the test response consists of checking words or symbols at particular places on the blank, the correcting can be greatly facilitated by the use of a stencil. A sheet of transparent material such as celluloid is placed over a blank in order to mark the correct places on it with India ink. The stencil can then be aligned over a blank that is being corrected, making it easy to note whether the marks on the blank correspond to those on the stencil. Instead of a transparent stencil one can be made of cardboard with holes cut at appropriate locations. The scorer merely looks for holes with symbols in them.

Various other minor points facilitate somewhat the scoring or statistical treatment of tests. It may be desirable to have the lines numbered, provided there is one test item to a line. If there are several items which he should mark in each line, the cumulative total from the beginning of the test to the end of the line in question may be indicated at the end of each line. This will save a few seconds in determining how many items have been attempted

after the correct or incorrect ones have been checked. Sometimes it is well to have printed at the end of each line the number of items that should have been marked in that line. In case it is undesirable for the subject to know this key, it may be concealed. For instance, if the test consists of numbers or letters that are to be canceled, the last number in the line may actually be the key that tells how many should have been marked or the last letter may represent the correct number in code. It sometimes facilitates matters to have the location of the answers staggered in two columns so that the odd-numbered answers are in one column and the even in another. This is of value where it is desired to total the two separately in order to check the reliability of the test by correlating one half of it with the other (cf. p. 165).

In some test projects it is convenient to have the answers on a separate pad rather than on the test blank itself. Even when using the blank the subject may transfer the answers to the margin. For example, in a test of the alternative answer form, he may underline the correct alternative and also put its number in a column at the right. This procedure may be carried a step further by having a separate pad with a column for each page of the test. When the subject works on the first page of the test he puts all his answers in the first column on the pad. If, for instance, in the first test question or item he thinks that the third alternative is correct, he enters "3" in the space at the top of the first column on the pad. His answer for the second item on this page will appear directly below the "3" that he has just entered. When he comes to the second page of the test he uses the second column of the pad in exactly the same way. By this method, with a test involving several hundred items the subject can enter all his answers on a 3 x 5 card. These answers can be corrected from a key which consists of a similar card or pad with the correct answers filled in. It is convenient to file and the original test blank can then be used repeatedly, thus saving expense.

Another time-saving device is to have the test pad prescored on the back. The intelligence tests used at the Ohio State University are handled in this fashion. Each item is of the multiple choice type with four or five alternatives. Small square "boxes" are numbered consecutively and the subject uses a phonograph needle mounted on a small wooden skewer to punch a hole in the

paper in the square corresponding to the correct alternative. The pad for the answers is not attached to the printed portion of the test. The latter has a full-sized envelope stapled to it with the opening to the right; a pad of the same size as the test is slipped into this envelope. The pad is divided into columns, the right column corresponding to the first page of the test, the next to the second page of the test, and so on. As the subject goes through the successive pages he pulls the pad out a little farther and the boxes for his answers are available just to the right of the printed material. He is able to juxtapose the correct materials by means of large numbers on the pages of the blank and at the head of the columns on the pad. Thus the answers for the entire test of seven pages and some 150 items appear on a single $8\frac{1}{2} \times 11$ " pad. The subject has punched holes through the boxes corresponding to his answers. The pad actually is double; and if the backing is removed, small squares are found already printed on the reverse of the page at the points where the punch marks should be if the responses are correct. The scorer merely counts the printed squares in which there is a hole. This speeds up the process of scoring tremendously. It also makes it possible to use the booklet over and over by inserting a new pad for each subject. Moreover, these pads are compact for filing.

Machine scoring constitutes another possibility. The answers are marked on a pad of appropriate size with a special pencil. Such a pad can be fed through a machine which gives the score immediately. The basic principle is that small electric brushes pass across the pad at various places and a pencil mark in the correct position closes a circuit across the two brushes because the pencil mark is of lower resistance than the paper. These changes of resistance are integrated in an electric circuit so that a meter will register. The meter can be calibrated to read the number of correct items instead of milliamperes. Scoring blanks in this way is extremely rapid but necessitates rental of the machine.

Small individual units have been devised whereby the subject punches keys when he is taking the test and the answer is recorded immediately by the device itself [10]. To administer group tests, however, it is necessary for each subject to have one of these units. The device is a compact metal box about 7 inches in each

dimension, with four keys protruding from the front. Inside is a drum which carries a strip of cardboard on its periphery. By means of a template, this strip has been punched to correspond to the correct answers; it resembles somewhat a record for a player piano. Small lugs drop through holes in the cardboard under certain conditions and engage mechanisms underneath. The subject reads a test item and if he thinks that the third alternative is correct, he presses key number three. The apparatus is so arranged by virtue of the cardboard inside that if the subject does press the correct key the mechanism steps along and he is ready for the second item. If, however, he presses the wrong key nothing happens, and he has to try another key. Meanwhile a window on the rear of the apparatus indicates the actual number of attempts, that is, the actual number of times a key has been pressed. After he has gone through the series of items one can read from the window the number of attempts the subject required to perform correctly the standard number of items. This procedure incidentally is useful in many teaching situations where the person can give himself a test and discover his mistakes as he goes along. It has not been used extensively in industrial testing, perhaps because it is a bit cumbersome to have a considerable number of such units and because with other procedures, such as those already described, the labor of scoring a group test can be greatly reduced.

Other tests have been arranged so that the subject must correct his own mistakes as he goes along. In a continuous choice reaction test where one of three lights appears and the subject has to press one of three keys corresponding to the lights, it can be so arranged by a magnetic ratchet stepping device that if he makes a mistake nothing happens, and he has to find the correct key before the lights change and give him the next one to which he is to respond. Thus the number of correct responses made in a given number of minutes is the final score as the mistakes have all been corrected and simply delayed the subject.

There may be instances where a momentary inspection of the test blank will suffice to indicate that it is not necessary to score it at all. If, for example, a critical score (p. 236, *infra*) has been set up—that is, a score below which no one will be hired—and if a glance at the number of items attempted in the test shows that it

is below the critical score, there is obviously no use in going any further. If a subject does not reach the critical limit even though all his answers are correct, it is not necessary to ascertain this latter fact in order to determine that he is unsuited for the job [3]. Incidentally in many tests there is a high correlation between actual score and number of attempts. With one self-administering intelligence test the correlation between the number of items attempted and the score on the whole test given with a time limit was .66. For one portion of the Minnesota Clerical Test a similar correlation was .93 and for another portion .92.

Scoring Speed and Accuracy. After a test has been corrected, one of the most serious problems is to determine what shall constitute the final score. The subject may omit some items; he may get others wrong. The omissions are not usually considered as serious a problem as the errors. Unless specific instructions have been given to omit no item and unless the subject has very patently skipped around and tried to pick the easiest ones, an occasional omission is overlooked and emphasis placed on those actually attempted. In cases where the test consists of finding certain things (as in canceling A's on a page) the omissions may be counted as errors or else an arbitrary formula devised to weight them. This problem, however, arises in only a limited number of tests, whereas the problem of speed and accuracy is present in a majority of mental tests. There are three ways in which the problem of speed and accuracy score may be handled. In the first place, errors may be neglected and speed alone or the number of items correct constitute the sole score. This is reasonably satisfactory when the errors are relatively few. In many kinds of tests the subjects will make comparatively few mistakes—if properly instructed, perhaps not over 5 per cent. If this is true of all the subjects, it is reasonably safe to neglect the errors. In some mechanical devices for test administration the subject is compelled to correct each error before proceeding to the next item (cf. p. 156).

In the second place, it is sometimes feasible to score only the accuracy or quality of the responses and to neglect the speed. This is to some extent true of the "power" tests (*supra*) in which a rather liberal time is given for the test or in which everyone finishes it and little account is taken of the time consumed. The

answers are then scored entirely on the basis of their quality or accuracy.

In the third place, speed and accuracy may be combined into a single score. This may be done either arbitrarily or statistically. The most usual practice is to penalize the subject a certain amount for each error and subtract the total penalty from the total number of correct items. In typewriting contests 5 words per error is the standard penalty. The examiner uses his judgment in determining the penalty; he decides whether mistakes are very serious in the particular situation for which the test is to be used, and if they are, he makes the penalty severe. In such cases it is wise to score a considerable number of blanks, using various degrees of penalty, and to study the results carefully to see the relative standing of the subjects with the different penalties.

In some types of test the appropriate weighting of these factors is rather obvious. If the subject chooses between two alternatives, there is an approximately even chance of getting the correct answer by guessing just as in tossing a coin there is an even chance of getting "heads." If there are 100 items and a subject knows absolutely nothing about them but simply marks them at random, he will get approximately 50 correct, while a subject who tries to work them out but goes slowly and painstakingly may not do more than 30, but get these 30 correct. The score of "number correct" will then be entirely misleading, for the first man ought to score zero. This situation is usually met by scoring the number right minus the number wrong. The argument is that the man who guesses on all of the items, and has 50 right and 50 wrong, will score 50 minus 50 or 0, while the man who does 30 correctly and makes no mistakes will receive 30 minus 0 or 30. This seems fair. Or suppose the second man actually knows 30 items but does not know the other 70, and guesses on them in addition to doing the 30 that he does know. He will then get about 65 correct—the 30 he knows plus 35 or half of those at which he guesses. He will likewise have 35 wrong—half of those at which he guesses. His score will be 65 minus 35 or 30, which is what he deserves for the 30 items he actually knows.

Although this method of scoring the two-alternative test is widely used, it has certain shortcomings. In the statement that if items are marked at random approximately 50 per cent will be

correct, the emphasis is on the "approximately." Many of us have encountered a situation where the pennies within a limited time did not match according to theoretical expectation. The laws of probability merely insure that if a large number of people guess at the items the people who get half correct will numerically exceed those who get any other number correct, but this former group is by no means the majority. It is quite similar to tossing coins. Suppose ten coins are tossed a large number of times. Five heads and five tails will be thrown in the long run more often than any other combination, but there is also a possibility of other results. Four heads, six heads, or even two or three heads may occur sometimes, although less frequently than the five heads. Exactly the same thing applies in guessing at test items where there are two alternatives just as there are two sides to a coin. It is possible to compute from the theory of probability what is to be expected in the long run.

Suppose that a test contains 10 items (such a brief test, of course, would not be used in the practical situation). It is possible to compute what percentage of the subjects who guess at the items will correctly guess 10, 9, 8, 7 items, etc. These percentages are given in the first part of Table 13. For instance, 0.1 of 1 per cent of the subjects will in the long run get all 10 items correct, 1 per cent will get 9 of them correct, etc. Similarly, if the test comprises a more reasonable number of items such as 50 (cf. the second part of the table), 0.1 of 1 per cent of the subjects will get 36 of them correct, 0.2 of 1 per cent will get 35 correct, etc. There are still smaller percentages which do not appear in the table for more than 36 or less than 14 items. In both instances it is to be noted that more subjects are apt to get just half of the items correct than are apt to make any other score. However, these subjects by no means constitute the majority—in the first instance they are about 25 per cent and in the second 11 per cent of the group. Obviously, if the test is scored according to the number correct, some subjects by mere guessing will get a fairly high score. Some allowance for this must be made. If the usual allowance is made by scoring the number right minus the number wrong, as indicated by the second column in each section of the table, and all the scores are called zero that are negative by the computation, this improves matters considerably. More than half

TABLE 13. PROBABLE PERCENTAGES OF INDIVIDUALS GUESSING CORRECTLY
VARIOUS NUMBERS OF ITEMS IN TESTS OF TWO-ALTERNATIVE FORM

<i>Test of 10 Items</i>			<i>Test of 50 Items</i>		
Number of Items Correct	Number Right Minus Number Wrong	Per Cent of Indi- viduals	Number of Items Correct	Number Right Minus Number Wrong	Per Cent of Indi- viduals
10	10	0.1	36	22	0.1
			35	20	0.2
9	8	1.0	34	18	0.4
			33	16	0.8
8	6	4.4	32	14	1.6
			31	12	2.7
7	4	11.7	30	10	4.2
			29	8	6.0
6	2	20.5	28	6	7.9
			27	4	9.6
5	0	24.6	26	2	10.8
			25	0	11.3
4	0	20.5	24	0	10.8
			23	0	9.6
3	0	11.7	22	0	7.9
			21	0	6.0
2	0	4.4	20	0	4.2
			19	0	2.7
1	0	1.0	18	0	1.6
			17	0	0.8
0	0	0.1	16	0	0.4
			15	0	0.2
			14	0	0.1

the group thus receive their deserved score of zero, but even then there are some who make rather high scores. In the 10-item test, for instance, 4 per cent of the individuals will make a score of 6, whereas they know nothing about the items; in the 50-item test 4 per cent will score 10 points. This method of scoring still tends to give some persons a higher score than they deserve.

If the subject marks the items of which he is certain and guesses at the others, chance, as just indicated in Table 13, may work to

his advantage or disadvantage. The usual procedure of requiring the subject to mark all the items, i.e., to guess when he is uncertain, and then scoring the result right minus wrong is about the best that can be done. But it has the shortcomings indicated. There is a considerable trend toward devising test items with 5 or 6 alternatives; this minimizes the probability of accidental success and obviates many of the difficulties just described.

The best scheme for combining speed and accuracy involves their evaluation statistically with reference to the situation in which the test is to be used. Suppose the test is devised for predicting ability in clerical work, and is to be evaluated by comparison of test scores with an occupational criterion, such as production figures or estimates of the office managers. Given this criterion, it is possible to correlate with it speed in the test and also accuracy in the test. It can then be determined which is the more closely correlated with clerical ability or which is the more valuable in predicting it. Moreover, it is possible by the technique of partial correlation to determine the best weighting for these two factors. This technique has already been mentioned (p. 60, *supra*) and will be discussed more fully in Chapter IX.

Not only are speed and accuracy related in some degree to the criterion, but they are related, perhaps inversely, to each other. It is necessary to determine what the relation of each to the criterion would be if the other were eliminated or kept constant. For instance, if a number of subjects could be obtained who had all exactly the same speed, it could be determined to what extent accuracy correlated with proficiency in clerical work for this limited group; and if another group could be found all with the same accuracy, the correlation of their speed with the criterion could be computed. It is seldom possible to find such groups, but it is possible, by the mathematical technique above mentioned, to obtain the same result from the actually available data. When these partial correlations are found—i.e., the intrinsic relation of speed and of accuracy to the criterion with the other factor constant—it can be determined exactly how much importance or weight should be attached to each. When the speed and accuracy are weighted according to this procedure, the combined scores will correlate more highly with the criterion than if they are weighted in any other fashion. This can be shown

theoretically or empirically. In an actual case the correlation coefficient between speed in the test—i.e., number of items completed—and the criterion is .60; the correlation between accuracy in the test—i.e., number of mistakes—and the criterion is -.50; and the correlation between speed and accuracy is -.20. This indicates that those who accomplish the largest number of test items tend to be most effective in the job, and vice versa, that those who make the fewest mistakes are likewise most effective in the job, and those who do the greatest number of items tend to make somewhat fewer mistakes, although this last relation is not very marked. Application of partial correlation technique indicates that the best scoring formula is:

$$\text{Criterion} = \text{Number right} - .76 \times \text{Number wrong.}$$

In other words, if each correct item counts 1 point, the subject should be penalized .76 of a point for each mistake. If the individual blanks are now scored by this formula with speed and accuracy weighted in this fashion, these weighted scores correlate with the criterion to the extent of .71, which is considerably better than the correlation of .60 that was obtained with speed alone. Hence weighting the two variables in this way materially improves the prediction of the criterion on the basis of the test scores.

In fact, for a problem like the present one it is not necessary to compute the partial correlations by the usual procedure. It is feasible to employ an equation actually derived from partial correlation theory and compute the weight that should be applied to the number of mistakes [13]. The formula is:

$$\text{Weighted score} = R + CW$$

in which R is the number right, W the number wrong, and C a constant. The formula for C is as follows:

$$\frac{\sigma_R (r_{IR} r_{RW} - r_{IW})}{\sigma_W (r_{IW} r_{RW} - r_{IR})}$$

σ_R means the standard deviation of the right scores; r_{IR} means the correlation between criterion (I) and number right; r_{RW} means the correlation between number right and number wrong, etc. It can be shown that if the weight is derived in this fashion and

the results are scored accordingly, the correlation between the test and the criterion will be higher than if the errors are weighted in any other way.

RELIABILITY OF TESTS

Repetition of Test. The test score is only an approximate measure of the ability in question. It is impossible for an ordinary test to be so devised that it will conform absolutely to all of the principles described in the preceding part of the chapter. Slight differences—in difficulty of items, for instance—are practically unavoidable. If a person makes a certain score in intelligence, it is somewhat inaccurate to say that this is his real score. Suppose that, instead of comprising 50 items, the test comprises a million. The latter test will probably give a more typical picture of the individual's ability. The results of the former set of items may deviate appreciably from the results of the latter. It is a question of how reliable a sample of the particular mental performance is embodied in the brief test. Or suppose that the same subjects take the same test over again. It may be found that their second scores are appreciably different from their initial scores. The crucial point, however, is whether in the second test the subjects maintain approximately their initial relative standing; in other words, whether a subject who makes a good score in the first test does likewise in the second, and vice versa.

The problem is analogous to that of ascertaining the reliability of some physical instrument. Suppose with a steel tape we measure the length of some objects such as a table and a desk and find them 60 and 58 inches respectively. Then we measure them again with the same tape and secure nearly the same results. If, on the other hand, we use a cloth tape we may find the desk longer than the table on the second occasion simply because we stretched the tape more when measuring the desk. Thus we may say that the steel tape is reliable and the cloth tape unreliable. By the same token we measure a considerable number of people with a mental test on two occasions and ascertain whether the person who does well on the first occasion does well on the second; i.e., we correlate the first and second administration of the test. If the correlation is high, say upward of .80, we feel that the test is fairly reliable.

This method of computing reliability may be vitiated by the effect of practice. Some subjects may profit by the practice more than others and thus suggest unreliability of the test, whereas the result may reflect merely the difference between the subjects in learning by experience. This in itself might be interesting, but it could better be investigated directly. This difficulty is especially marked with tests where a subject can memorize certain items. For routine performance, such as canceling symbols on a page, repetition of the test may be an adequate method for determining reliability.

Another difficulty with the test retest method may be noted. The subjects may change many items but go in both directions—that is, change some that were incorrect so that they are now correct, and vice versa—so that the total scores are about the same. Thus technically the test would show high reliability although the subjects were not very consistent about their performance. In cases where this might be suspected it would be advisable to examine the individual items for changed responses. In a test of tact in which the subject indicates what one should do in certain situations, the reliability on the basis of test retest was .74. However, out of the 42 items, the average subject changed about 10 when selecting the best course of action, and between 11 and 12 when selecting the worst [4].

Correlating Two Forms of Test. Instead of actually repeating the same test it may be prepared in two parallel forms which are presumably comparable in difficulty. As suggested earlier, when test items are made up and their difficulty is measured, usually many more are evaluated than are necessary for a single test. It is thus possible to construct two forms of about the same difficulty by pairing each item in the first form with a similar one in the second. Correlating one form with the other to determine reliability obviates the difficulties caused by practice.

Split-half Method. Instead of constructing tests in two forms or giving the same test twice, one form may be given but divided into two parts for the computation of reliability. If a test comprises 100 items the first half may be correlated with the last half. This would be feasible in a work limit test in which all the subjects attempted all the items. In a time limit test some subjects

would not get as far along in the second half as would others and comparison between the two halves would be meaningless.

A common practice is to number the items and constitute one half of the test from the odd-numbered items and the other half from the even-numbered items. This computation may be facilitated by having the places for the answers on the test blank staggered so that the odd ones will be in one column and the even in another. One possible shortcoming of this procedure is that it gives no notion of day-to-day reliability, i.e., whether the test is seriously affected by daily condition, mood, and the like. However, the other advantages of this procedure outweigh this difficulty so that it is widely used. With this split-half method we are actually computing the reliability of half a test. If the test has 100 items we correlate the first, third, fifth, and so on, with the second, fourth, and sixth. We are simply correlating 50 items with 50 items and getting the reliability of a 50-item test. A simple formula is available that corrects this reliability coefficient to give what it would have been if we had correlated 100 items with another 100. The formula is:

$$\frac{2r}{1+r}$$

where r is the correlation between the odd- and even-numbered items. Incidentally, by a slight extension of this formula we can determine how the reliability of any test can be raised by increasing its length by any designated amount. On occasion we may be dissatisfied with the reliability of a test and wish to increase it by making the test longer.

Causes of Unreliability. Some studies have been made of the factors which contribute to reliability or unreliability of tests. Some of these are reported by Symonds [12] and also summarized by Guilford [6]. A few of the main factors may be noted. The length of the test was just mentioned. The narrower the range of difficulty of the items the greater the reliability because very difficult items add nothing. Interdependent items reduce the reliability. Such items are passed or failed together and this is about the same thing as reducing the length of the test. More objective scoring promotes greater reliability. A large element of chance in answering decreases reliability. For example, with a true and

false test the reliability was .84. When the number of alternative responses was seven, this same test had a reliability of .91. As the material is more homogeneous the reliability increases. If the subjects have had more common experience, the reliability will be greater because they have been exposed to the same things. Anything leading to misinterpretation of an item is undesirable, such as emotional coloring of words or poor sentence construction. Variations in set such as speed vs. accuracy should be avoided. Variation in incentive is another undesirable factor. Perseveration or carry-over of some emotional experience or some distraction may produce unreliability. Accidents such as the breaking of a pencil, or illness, worry, or excitement may be factors. Cheating or coaching are possible elements.

It is thus essential after a test has been devised to investigate its reliability before putting it into practical use. In dealing with different groups of employees or applicants, we want always to apply the same mental measurement, just as in determining the dimensions of different house lots we prefer to use a tape that is always 100 feet long. If an unreliable test seems to indicate vocational aptitude with one group of employees, it may utterly fail to have any prognostic value with another group.

SUMMARY

Mental tests such as are used for predicting vocational aptitude are devices for measuring a typical sample of mental or motor performance. Their administration may involve the examination of one individual at a time or of a group of subjects simultaneously. The latter procedure saves much time, although there is more chance for the subject to fail to follow directions and there is less opportunity for clinical observation of his extraneous reactions. With a small number of subjects, however, and ample floor space, the group test has most of the advantages of the individual test. The subject's response may be oral, it may be written on a blank, or it may involve some performance with implements or apparatus. The type of response may depend on the wording of the question, the location of the answer, the selection of the answer from alternatives, or matching. Tests are given with either a time limit or a work limit. The former is generally used in group tests; the latter cannot be used unless the subjects can be trusted

to record their own times. It is not feasible to have the examiner record the individual subjects' times in the work limit method unless these times are relatively long, as in the "omnibus" test. The time limit must be sufficiently short so that no subjects will quite finish. Otherwise it will be impossible to differentiate between the proficiency of those who complete the test. Work and time limits yield similar results. Certain general precautions must be observed in test procedure, such as maintaining standard conditions, having the subject in a cooperative attitude, and having the examiner perfectly familiar with the technique so that the test will go smoothly.

In selecting test material all items of a given kind may be of approximately equal difficulty with the emphasis placed on speed, or they may increase in difficulty with the emphasis on ultimate level of attainment or "power." The material is usually arranged to group together items of a given sort so as to facilitate separate scoring, but sometimes the different kinds of items are intermixed. This "omnibus" form of test may present all items of a given type of approximately equal difficulty (cycle omnibus) or the items may increase in difficulty throughout the test (spiral omnibus). In preparing a test, alternative material of the same difficulty as the original must be provided to meet the situation if blanks reach the hands of subjects before they are tested. The test must be sensitive, i.e., give a wide range of scores. This can be accomplished by having a considerable number of items in the test and having it of appropriate difficulty, neither extremely easy nor very hard, for the group taking it.

Test instructions must be kept absolutely standard and constant whenever the test is used. They should be sufficiently clear to enable the subjects to understand perfectly what is wanted. It is well to come down to the lowest intellectual level present in the group. Instructions usually comprise explanation, illustration, and practice. It is important to determine how ability in a test is influenced by practice and to decide how much practice should be given before the actual performance on which the subject is scored. Printed instructions are usually preferable to oral because of their more rigidly standard character. Incentive while taking tests must be kept at a maximum in order to keep it con-

stant. This may be done by the instructions or by utilizing various motives such as pride, cooperation, or competition.

The scoring of tests must be unequivocal so that different persons scoring the same subject's blank will obtain identical results. The blank may be arranged with a view to ease of scoring by means of stencils, by the location of the answers in convenient fashion, or by a separate answer pad. Other possibilities are to have the pad prescored on the back or to use a machine which integrates the pencil marks electrically. In obtaining a final score for a test, the question of the relative importance of speed and accuracy arises. Sometimes speed is stressed and accuracy neglected, sometimes the reverse, and sometimes both are combined into a single score. This may be done purely arbitrarily or, in some instances, by considering the probability of getting correct answers by guessing. The best procedure is to correlate speed and accuracy separately with the vocational or other criterion and weight them by the partial correlation technique. Scores combined in this manner will give a more valid prediction of the criterion than those combined in any other way.

Before a test is put into practical use its reliability should be determined. This may be done by giving it twice, by administering two forms of the test, or by splitting the single test into two halves, and correlating the two scores obtained by any of these procedures. If these correlations are high, the test may then be used with impunity in employment research.

REFERENCES

1. Atwell, E. R., and Wells, F. L. Wide-Range Multiple Choice Vocabulary Test. *Journal of Applied Psychology*, 1937, 21, 550-555.
2. Bingham, W. V. *Aptitude Testing*. New York, Harper, 1937, 390 pp.
3. Copeland, H. A. Some Characteristics of Three Tests Used to Predict Clerical Success. *Journal of Applied Psychology*, 1936, 32, 461-470.
4. Gibbons, C. C., and Porter, J. P. Some Aspects of Social Adaptability Among Adolescents. *Journal of Applied Psychology*, 1939, 23, 508-520.
5. Greene, E. B. Practice Effects on Various Types of Standard Tests. *American Journal of Psychology*, 1937, 49, 67-75.

6. Guilford, J. P. *Psychometric Methods*. New York, McGraw-Hill, 1936, 566 pp.
7. Langdon, J. N. An Experimental Study of Certain Forms of Manual Dexterity. Industrial Health Research Board, *Report No. 66*, 1937, 64 pp.
8. O'Connor, J., and Filley, M. E. A Junior English Vocabulary Test. *Personnel Journal*, 1933, 12, 204-212.
9. Paterson, D. G., and Tinker, M. A. Time Limit vs. Work Limit Methods. *American Journal of Psychology*, 1930, 42, 101-104.
10. Pressey, S. L. A Machine for Automatic Teaching of Drill Material. *School and Society*, 1927, 25, 549-552.
11. Reed, H. B. The Influence of Practice on Individual Differences. *School and Society*, 1931, 34, 100-102.
12. Symonds, P. M. Factors Influencing Test Reliability. *Journal of Educational Psychology*, 1928, 19, 73-87.
13. Thurstone, L. L. A Scoring Method for Mental Tests. *Psychological Bulletin*, 1919, 16, 235-240.
14. Thurstone, L. L. Mental Tests for Engineering Students. *Engineering Education*, 1923, 13, 263-318.
15. Viteles, M. S. The Influence of Training on Motor Test Performance. *Journal of Experimental Psychology*, 1933, 16, 556-564.
16. Workman, B. E. The Work Limit Method vs. the Time Limit Method in the Instruction and Administration of College Ability Tests. *Ohio College Association Bulletin No. 79*, 1932, pp. 963-973 (mimeographed).

Chapter VI

THE CRITERION

NECESSITY

Basis for Evaluating Tests. Mental tests, like other instruments, do not always serve the purpose for which they are designed. A radio transmitter may have ample power and a good antenna and yet be unable to put a satisfactory signal into Australia. A mental test may be reliable, objective, and fool-proof, and still fail utterly to separate the sheep from the goats in the stitching room. The psychologist is no more omniscient than the electrical engineer. In either case it is necessary to give the instrument an actual trial and see if it does what it is supposed to do. Consequently, before psychological tests can be validly used for employment purposes they must themselves be tested by comparing, in a typical group of workers, efficiency in the tests with efficiency in the job. This implies two measures for each person on whom the tests are standardized—his test score and some figure that represents his occupational efficiency. This latter—the thing by which the tests are actually evaluated and the thing which it is desired ultimately to be able to predict—is technically called the *criterion*.

The need, however, is not merely for a criterion as such, but for one that is as reliable and accurate as possible. The value of the entire project depends upon it, because it is the standard used in evaluating the tests. If the criterion is inaccurate, the tests designed to predict it will be proportionately inaccurate. In an actual instance where the criterion consisted of estimates by the foreman, this particular individual had a bias in favor of the older employees. What he turned in was essentially a

ranking on the basis of age. The research staff took this criterion at its face value and set about to develop tests which would correlate with it. After considerable research they had a battery which did show a respectable validity. The result of all this was that they could now give these tests to a group of applicants and make a pretty accurate prediction as to the applicant's age—information which could, of course, have been obtained much more expeditiously by other means. Similarly if production figures used as a criterion are based on piecework rates that have been unscientifically and carelessly determined, the tests will not predict proficiency in that work. For the best that the tests can do is to predict the criterion by which they are evaluated. If the criterion is inadequate, the entire project is resting on shifting sands. Hence every effort should be made to get the best possible data regarding the workers' ability in the job and to handle these data in the best scientific fashion once they are obtained.

Insuring Availability of the Criterion. Before undertaking a project of this sort, it is well to make certain that the criterion will be available when needed. One should ascertain whether production records are kept in such a form that they can be utilized or whether foremen and other supervisory executives are willing to cooperate in making ratings. If the tests are given to applicants for employment rather than present employees, it is well to initiate at the outset some procedure for following up those tested (cf. p. 280). As a matter of fact, it is a good policy, when possible, to obtain the criterion in advance of any testing at all. If many employees are tested with the understanding that subsequently the foreman will rate them and the foreman dies, the efforts will have been largely wasted. One of the committees that approached the problem of tests for aviators in 1917 gave a considerable range of tests to a large number of cadets at one of the ground schools; the understanding was that these men would be sent to a flying field from which a subsequent record of their progress could be obtained. Many of them, however, were sent directly to France for their flying instruction so that it was impossible to obtain the criterion. A few experiences of this sort impress the psychologist with the importance of making certain of the criterion in advance or at least insuring its ultimate availability before undertaking any testing project.

Types of Criteria. Criteria may be classified as objective and subjective. The former involves actual daily production on the job or sometimes a work sample, that is, a standardized sample of the job done under observation which can be graded as to quantity or quality. The subjective criterion involves direct judgment of the individual and his performance. For this purpose the judge must actually know him and have seen him in action. If several criteria are available in a given case, there then arises the problem of reducing them to common terms and combining them into a single figure. These topics will be discussed in order.

PRODUCTION FIGURES

Production is after all the most obvious criterion. It is the thing which the management is ultimately interested in predicting and under favorable conditions is probably the best indication of a man's ability in the job. In many instances production figures are comparatively easy to obtain because production records are kept for purposes of making out the payroll. In operations like checking, pasting, assembling, and making various parts of shoes, garments, or tires, a record of the number of pieces done per unit time is often available. Some machines such as looms carry automatic counters that record the number of operations performed. Sales records are frequently available for persons in the marketing end of industry. Records may be kept of the number of pounds of mail distributed by a postal clerk per unit time or the average daily revenue of taxi drivers. Even in evaluating a foreman's efficiency the production of his department may be significant. It goes without saying that an adequate sample of production should be secured. Common practice is to determine the production per hour based on records for several hundred hours.

Rather wide ranges of ability as indicated by production are often discovered under industrial conditions. Operators punching Hollerith cards vary from 28 to 133 cards per hour. The fastest operator is thus almost five times as efficient as the slowest. In various operations in shoe manufacture, the ratios of best to worst vary from 1.5 to 2. Among coal miners the best produce almost 12 times as much as the worst. An extreme case is insurance solicitors; in this field men with equal experience vary to

such an extent that one man sells 350 times as much as another. It is these wide differences in production that it is desired to predict.

Workers' Attitude. There are, however, a number of points that should be scrutinized in a given situation before the production criterion is accepted as adequate. In the first place, the attitude of the worker toward his work must be considered. His production record is not a true measure of his ability in the job unless he devotes his best effort to the job. This implies that he is industrious rather than lazy, that he is not sick or worried or engrossed in other matters, and that he has ample incentive to bring out his maximum endeavor. Some of these contingencies it may be impossible to ascertain, but frequently the foreman or supervisor will have a more or less personal touch with the men and be able to supply this information. In obtaining estimates by foremen, such points as the above will sometimes be covered if they think them important.

As to actual incentive, this is to an appreciable extent assured in the case of pieceworkers. These workers are paid so much per unit of work—i.e., their wage depends directly on what they do—and hence in most cases they will do their best in order to have a large pay envelope. Even in the case of pieceworkers, however, the foreman's judgment is by no means unnecessary, for there are instances of "stereotyping of output" in which, in spite of the possibility of more pay, men voluntarily limit their production. It appears that a worker will exert himself up to what seems a reasonable limit; beyond this the discomfort and inconvenience involved in doing more prevent him from reaching a higher level of proficiency. There is a balance between the worth-whileness and the exertion of the job. Special cases like the following may be encountered. Adolescent girls in a garment manufacturing shop threaded needles for the more skilled workers and were paid on a piece-rate basis. The installation of a bonus system did not increase production as had been anticipated. It developed subsequently that because these girls were minors their parents made them bring their pay envelopes home unopened; the girls received only an "allowance" by their parents. Hence, if they did more work it was the parents who received the bonus. When the girls were assigned a quota considerably

above the average level of performance and told that after they had finished the quota they could go home for the day, most of them were through at 2:00 P. M. In cases like this, obviously the production records would constitute a very poor criterion.

This question of attitude is usually a much more serious problem in the case of day work in which the person is paid a flat time rate regardless of the amount of work done. It may be necessary for him merely to keep moving in order to hold the job. Often no official record is kept of his performance, although it is sometimes possible to collate such figures from time slips. Where the time rate is flexible it might seem that a man's rate would indirectly reflect production, but it is just about as apt to reflect his length of service, his aggressiveness in asking for a raise, the size of his family, or his consanguinity with the foreman. Sometimes records are kept with a view to determining when to promote the worker or to raise the time rate. In such instances the incentive is partially obtained. But the production of day workers is at best a precarious criterion.

Equivalent Units of Production. Another problem is whether the units of production used in determining the scores of different men in the same occupation are equivalent. If all the workers involved in a given study are doing exactly the same job—e.g., all making a 6-inch tire or attaching number 3 labels or typewriting form letter number 5—the actual numbers of pieces done per hour by different individuals are of course satisfactory units by which to compare those individuals. If, however, one is building a 6-inch tire and another a 7-inch, if one is pasting small labels and another large, if one is typewriting short letters and another long letters, it is not fair to compare them in terms of tires or labels or letters. Likewise, a salesman's production may be influenced by the difficulty of the territory, by competition, by prejudice, or by whether he sells furniture or notions. In some such cases it may be possible to divide the work into smaller comparable units, such as number of lines typewritten or even number of strokes made, if appropriate recording devices are attached to the machine. In other cases recourse may be had to the results of the time study that has been made in setting piecework rates. If the rate has been set so that a piece that takes twice as long as another is given twice the pay, then the pay

per hour is a reliable index of production. In other instances it may be better to take production as a percentage of the standard set by time study. If, for instance, the standard is 50 units per hour and the worker does 60 units, his score is 120 per cent, whereas if he does only 40, his score is 80 per cent. For operators of power sewing machines, for example, there were available for each worker "earned hours," that is, the time set by the company for completing the given unit of work. This figure was based on time and motion studies. There were also available the "clock hours" which the worker actually required in order to complete the unit. The former was divided by the latter to obtain an index of efficiency. If the time study has been properly made, this seems to be a fair way to secure comparative production figures.

In operations where persons work together as a team—for example, in building tires or folding tablecloths—from the standpoint of one worker's production the team-mate constitutes an extraneous factor. The slower worker of the pair may set the pace, or the faster one may do so providing the inferior one can keep up. When the same individual worked with two teams folding tablecloths in a laundry, the team production in one case was 35 per cent better than in the other [3, 168]. In some operations it may not be a question of two workers collaborating simultaneously but rather operating in sequence. The production of the second one is contingent upon that of the first and would not be an appropriate criterion for the second. Lack of materials is a similar variable that might invalidate production criteria. At any rate, obtaining the criterion is more difficult under such circumstances.

Experience. Varied amounts of experience on the part of the workers may invalidate the results. Those who have not been at the job long enough to reach their maximum efficiency naturally will not have a record that is typical of what their innate ability will enable them to do. All such cases should be discovered if possible, and allowance made or their results excluded. They may be located frequently on the basis of the foreman's judgment. Sometimes, in a given job, a study of the records of new workers can establish about how long a time is required, on the average, before maximum efficiency is attained. The results of men who have not worked at the job this length of time may

be excluded. Or it may be possible, on the basis of the record of the individual in question over a considerable time, to determine whether he is still improving or has reached his maximum. At any rate, attention must be given to this factor of experience. If workers are used who have had time to reach their maximum efficiency and if records of piecework production are accumulated over several weeks and reduced to pieces per hour, a satisfactory criterion will generally be obtained.

Combinations of Production Data. Sometimes the objective criterion involves production data of various sorts which must be combined in some way. This was true in an investigation of retail selling [5, 80]. Some of the items contributing to a criterion for this type of work were number of sales; average amount of purchases; actual credit; total credit of the individual plus the department credit; total net sales, that is, gross sales minus total credit; daily sales quota, namely, the salary divided by the average selling cost for the department; number of days of selling, that is, the number of days worked; the actual quota; any bonus obtained for selling more than the quota; extra selling costs, that is, special bonus for selling slow-moving items; cost per cent, that is, the total salary divided by the total net sales; clerical errors; handwriting errors; errors in computation; grades by the service shopper who went around incognito purchasing from the sales people; a sales rating based on the past few months; and finally actual salary. These items, which with a few exceptions were objective, were combined into a final total to constitute the actual criterion used. In using production data for a criterion the problem may arise of the comparative importance of speed and accuracy or quantity and quality. It may be feasible to combine the two into a single score as the production criterion. For instance, in punching Hollerith cards it was found that 13.7 cards could be punched while one error was being corrected. Consequently in determining the production of the operators 13.7 cards were subtracted from the score for every error made.

WORK SAMPLES

Rather than using day-to-day production as a criterion, in some cases it is preferable to employ work samples. These utilize

per hour is a reliable index of production. In other instances it may be better to take production as a percentage of the standard set by time study. If, for instance, the standard is 50 units per hour and the worker does 60 units, his score is 120 per cent, whereas if he does only 40, his score is 80 per cent. For operators of power sewing machines, for example, there were available for each worker "earned hours," that is, the time set by the company for completing the given unit of work. This figure was based on time and motion studies. There were also available the "clock hours" which the worker actually required in order to complete the unit. The former was divided by the latter to obtain an index of efficiency. If the time study has been properly made, this seems to be a fair way to secure comparative production figures.

In operations where persons work together as a team—for example, in building tires or folding tablecloths—from the standpoint of one worker's production the team-mate constitutes an extraneous factor. The slower worker of the pair may set the pace, or the faster one may do so providing the inferior one can keep up. When the same individual worked with two teams folding tablecloths in a laundry, the team production in one case was 35 per cent better than in the other [3, 168]. In some operations it may not be a question of two workers collaborating simultaneously but rather operating in sequence. The production of the second one is contingent upon that of the first and would not be an appropriate criterion for the second. Lack of materials is a similar variable that might invalidate production criteria. At any rate, obtaining the criterion is more difficult under such circumstances.

Experience. Varied amounts of experience on the part of the workers may invalidate the results. Those who have not been at the job long enough to reach their maximum efficiency naturally will not have a record that is typical of what their innate ability will enable them to do. All such cases should be discovered if possible, and allowance made or their results excluded. They may be located frequently on the basis of the foreman's judgment. Sometimes, in a given job, a study of the records of new workers can establish about how long a time is required, on the average, before maximum efficiency is attained. The results of men who have not worked at the job this length of time may

be excluded. Or it may be possible, on the basis of the record of the individual in question over a considerable time, to determine whether he is still improving or has reached his maximum. At any rate, attention must be given to this factor of experience. If workers are used who have had time to reach their maximum efficiency and if records of piecework production are accumulated over several weeks and reduced to pieces per hour, a satisfactory criterion will generally be obtained.

Combinations of Production Data. Sometimes the objective criterion involves production data of various sorts which must be combined in some way. This was true in an investigation of retail selling [5, 80]. Some of the items contributing to a criterion for this type of work were number of sales; average amount of purchases; actual credit; total credit of the individual plus the department credit; total net sales, that is, gross sales minus total credit; daily sales quota, namely, the salary divided by the average selling cost for the department; number of days of selling, that is, the number of days worked; the actual quota; any bonus obtained for selling more than the quota; extra selling costs, that is, special bonus for selling slow-moving items; cost per cent, that is, the total salary divided by the total net sales; clerical errors; handwriting errors; errors in computation; grades by the service shopper who went around incognito purchasing from the sales people; a sales rating based on the past few months; and finally actual salary. These items, which with a few exceptions were objective, were combined into a final total to constitute the actual criterion used. In using production data for a criterion the problem may arise of the comparative importance of speed and accuracy or quantity and quality. It may be feasible to combine the two into a single score as the production criterion. For instance, in punching Hollerith cards it was found that 13.7 cards could be punched while one error was being corrected. Consequently in determining the production of the operators 13.7 cards were subtracted from the score for every error made.

WORK SAMPLES

Rather than using day-to-day production as a criterion, in some cases it is preferable to employ work samples. These utilize

certain products made by the worker which can be evaluated by persons who do not know him and without his being present. In evaluating the Minnesota Mechanical Abilities Tests, students in a sheet metal shop made a foot scraper, a cookie cutter, a rectangular box, a dustpan, and a funnel. These products were graded by expert judges and the grades used as a criterion [4, 147]. Similar procedures were followed in printing and wood-working. The subjects set standard pages of type or made standard products such as a ruler, a gameboard, or a towel rack, and these were evaluated by expert judges. Similar work samples have been used by the U.S. Employment Service for adding-machine operators, calculating-machine operators and card-punch-machine operators.

In work samples it is essential that the sample be typical. With calculating-machine operators a sample of addition alone would be inadequate because the operators have to do a great deal of multiplication. Hence it is necessary to do something in the way of job analysis in order to determine which aspects of the work are the most typical, and to devise the sample accordingly. Some question may be raised as to classifying the work sample as an objective criterion because it necessitates judgments of experts regarding the products and such judgments are necessarily subjective. However, the judgments are subdivided to include various detailed aspects of the product, with a detailed quantitative rating for each aspect. After all, the important consideration is the reliability of the criterion rather than its classification.

Reliability of Objective Criteria. In dealing with production figures effort should be made to determine the reliability of the data. This is the same kind of problem as that encountered with test scores (cf. p. 164). If production per hour is computed for one week it may likewise be computed for another similar period and the two measures correlated. If workmen who have relatively high production per hour during one week have a similar high production during another week, and vice versa—i.e., if the correlation between the records for the two weeks is large—this production criterion may be considered reliable. Similarly with work samples, results on two of the items made by the subjects—the cookie cutter and the funnel—may be correlated, or the

evaluation of the product by one expert may be correlated with that by another expert. In actual practice reliable production criteria are found more often than are reliable estimates by foremen. This is probably due to the fact that the former data are more objective than the latter and do not involve personal idiosyncrasies on the part of the foreman making the estimate. More reliable estimates of a person's height could be made with an objective yardstick than by combining the subjective judgments of his acquaintances. If it were always possible to obtain the production record under ideal conditions conforming to the various factors outlined above, estimates by foremen probably could be dispensed with. Unfortunately this is seldom the case.

ESTIMATES BY SUPERIORS

Estimates by an employee's superiors are almost always available as criteria. Practically every member of a concern's personnel is "under" somebody else. There is someone who exercises a certain amount of supervision over him and who has some notion as to the kind of work he is doing and his value to the company. If this superior has watched the man as closely as he ought, is willing to make estimates, and is sufficiently careful in making them, they are of some value. The type of estimates to be discussed in this connection should be distinguished, however, from the systematic rating scales to be presented in a subsequent chapter. These latter involve the separate judgment of a considerable number of traits that are not measurable by mental tests, and the ratings are used in lieu of test procedure. In the present connection the estimate usually is of only one thing, an overall judgment such as "efficiency in the job," and it is an estimate of something that it is hoped to predict by means of tests. The estimates that form the basis of the criterion are not usually as complicated or as extensive as those involved in rating scales.

Suppose that one or more foremen¹ are going to make esti-

¹ In the following discussion, estimates of foremen will be mentioned for the most part. While perhaps the majority of instances encountered in actual practice involve estimates of industrial workers by their foremen, the same principles apply to office workers rated by their managers and supervisors, or to executives or salesmen rated by their superiors. While foremen's estimates are used for purposes of illustration, the methods described are of general application.

mates of a given group of workmen. There are several ways of proceeding to the actual process of judging. The men may be grouped into a number of classes, they may be arranged in order from best to worst, or they may be rated systematically on a linear scale.

Estimates by Grouping. The simplest and likewise the least reliable method of making these estimates is to divide the men into groups on the basis of their ability. Sometimes as few as two groups are used. The foreman is directed to divide his men into good and poor. He may do this by making out two lists himself or he may be given a list of all the men and be asked to check it with appropriate symbols. The difficulty with this procedure is that it assumes a dichotomy between good and poor, whereas ordinarily all degrees of ability are represented. Moreover, these data do not lend themselves to careful statistical treatment. The most that can be done is to compute the average test score made by each group and note whether the good workers exceed the poor workers in test scores, whereas it is highly desirable to use the procedure of correlation in order to be able to predict the *probability* of success in the job on the basis of the tests. Matters may be somewhat improved, if only two groups are to be used, by selecting smaller groups at the extremes of ability. If there are 100 men it may be better to select the 25 best and the 25 worst than to pick the 50 good and the 50 poor. This obviates to some extent the assumption of a dichotomy, although even within an extreme group there are doubtless marked differences in ability. The average test scores made by the two extreme groups probably will differ more than the average scores made by the two groups that comprise all the men—providing the tests are of any value at all. It may be possible to assign some arbitrary value to each group and compute a rough correlation coefficient which will be meaningful.

In making estimates by grouping, it is desirable, however, to have more than two groups. The foremen may be directed to divide the men into three groups—good, average, and poor. A five-group arrangement is quite common. One class is labeled “outstanding, among the best workers in the department”; the second, “superior, above average but not outstanding”; the third is “average, neither superior nor inferior”; the fourth, “below aver-

age but not poor"; and the fifth, "among the poorest workmen in the department." In general, the more groups the better, up to a certain limit; because ability in the job is actually a continuous variable—i.e., there is a continuous gradation from worst to best—and the use of more groups gives a closer approach to such continuity. A classification into ten groups is fairly satisfactory because in correlation procedure the measures are often grouped into as few as ten classes. An important statistical consideration is involved with reference to the size of the groups. We know from the results of measurements of large numbers of human traits that there are usually more people of average ability than of any other degree, and that as we go up or down toward the extremes the numbers decrease so that there are very few with extremely good or extremely poor ability. This idea of the normal frequency curve is discussed below. This same thing holds true for most occupational abilities, and theoretically the foreman's estimates should comprise a large middle class with smaller classes above and below it, and still smaller classes above and below these. However, this concept is probably too complicated for the average foreman who is making the ratings; it is perhaps better either to neglect it or else to get the ratings in a quantitative form as described below and then make the proper divisions if this seems desirable. If the method of grouping is to be used at all, the safest rule is probably to use as many groups as possible (up to some reasonable limit such as twenty) and specify them by careful qualitative description.

Estimates by Ranking. A somewhat better procedure than the foregoing is the method of ranks or order of merit. It consists of arranging the individuals in order from best to worst. The names may be written on cards and the cards arranged in order, or the names may be in an alphabetical list and numbered.² This method is simple and lends itself readily to subsequent statistical treatment of the results, for it is possible to rank test

² It is possible to assign two or more persons the same rank if desired. In such cases, however, for statistical reasons they must each be assigned a rank obtained by averaging the ranks they would have received if they differed slightly. For instance, if they are tied for third and fourth place, they should both be ranked 3.5 and the next inferior should be 5. If three persons are tied for fifth, sixth, and seventh places, they should all be ranked 6 and the next person below them 8.

scores similarly and correlate the two sets of ranks. However, this method overlooks one thing. There is nothing to indicate whether the steps between successive pairs of ranks are equal or otherwise, and in handling the data the only possible procedure is to assume that they are equal. It must be assumed that the man ranked 1 is just as much superior to the man ranked 2 as the latter is to the man ranked 3. As a matter of fact, this may not be the case. Suppose that the actual values of the three best persons in occupational ability or test score or anything else are represented by the numbers 75, 60, and 59. In the rank method they will be marked 1, 2, and 3, and the assumption made that the difference between 75 and 60 is the same as that between 60 and 59. This assumption entirely obscures the comparatively great superiority of the first individual. Nevertheless, if there is a considerable number of men in the group that is being ranked, this assumption of equal steps will not make such a tremendous difference. Inasmuch as the method is simple and easily administered, it is widely used.

A variation of the ranking method or one which at least involves comparison between the individuals is the *paired comparison* method. If there are ten people in the group to be rated, instead of being ranked from one to ten each person is compared with every other one of the ten, i.e., 45 comparisons. The foreman is given these names in pairs and is asked to indicate in each case which of the two is the better workman. In the present example each workman has the opportunity to be chosen nine times in comparison with the others. Merely totaling the frequency with which each one is chosen gives an index of his comparative standing. This method is somewhat more systematic than the ranking procedure for comparing the workmen with each other.

Estimates on a Linear Scale. Linear scale estimates are usually the most desirable. A blank is provided on which the names of the men to be rated are typed at the left; each name is followed by a line of uniform length. The foreman makes a check mark at some point along this line to indicate his judgment. The right end of the line may indicate highest ability and the left end lowest ability. The farther to the right the mark is placed the better man it indicates. Inasmuch as the lines are of uniform length it is possible, after the ratings have been made, to con-

vert them into figures by measuring the distance of each check mark from the left.

While the linear scale may be presented in the above form with a mere indication of extremes, it is better to give some notion of the intermediate steps, and, although the ultimate figures take no account of them, to provide classes or other specifications as a guide to the person making the estimates. One method is to have the blank ruled into a number of columns. There may be three columns headed "poor," "average," and "good," either of uniform width or with the "average" column wider than the other two. A five-column arrangement has been found quite satisfactory although it has no theoretical superiority to other arrangements that might be devised. A portion of this blank appears as follows:

Name	Lowest Fifth	Next Lowest Fifth	Middle Fifth	Next Highest Fifth	Highest Fifth
Adams
Andrews
Briggs

It is well to arrange the width of the columns in some convenient unit. The blank is accompanied by directions such as the following: "Imagine all the men you have ever known who worked at this job divided into five classes with reference to their ability in the job—a highest fifth, a next highest fifth, a middle or average fifth, a next lowest fifth, and a lowest fifth. Put a cross somewhere along the line after each man's name to indicate in which group he belongs. Moreover, if he stands high in a group place the cross toward the right of the column, and if he stands low in that group, place the cross toward the left. In other words, the greater a man's ability in the job the farther to the right the cross is to be placed." This sort of explanation is usually intelligible to the average foreman, but it is well to discuss the matter with him and bring out any misunderstandings on his part. After a blank like the above has been marked, it is a simple matter to measure the distance of each mark from the left edge

of the left column in some convenient unit such as millimeters or fractions of an inch. This yields a quantitative expression of the ability in question. If the measures are to be grouped ultimately for a statistical treatment into, say, 10 or 15 classes of equal size, a transparent stencil ruled with 10 or 15 columns and numbered at the top may be placed over the blank. Each check mark is noted with reference to the column of the stencil in which it falls and the number at the top of that column is recorded as the criterion score.

Another scheme for indicating intermediate steps between the extremes of ability is to put descriptive adjectives along the line on which the rating is to be made. This is called the graphic scale technique. If, for instance, the criterion is to consist of an estimate of "quality of work" each line may appear thus:

Many errors	Careless	Good quality	Practically perfect workmanship
----------------	----------	-----------------	------------------------------------

These adjectives may be repeated with every line on the blank or may appear only at the top. This procedure is used more frequently in rating scale technique where a considerable number of traits are to be rated for each man. There such descriptions are more essential for guidance because the rater is considering one trait after another. This method will be discussed more at length in Chapter XII on rating scales.

One other point that applies to all the foregoing methods for securing estimates should be mentioned. An estimate of a worker before he has been at the job a sufficient length of time to reach his maximum proficiency is of little value. Many of us have been agreeably surprised at, or disillusioned by, the ultimate proficiency of an employee in contrast with our initial impression. Hence those who are making the estimates should consider whether the persons concerned have been at the job long enough to reach their ultimate level. If the foreman is not certain about a particular man and cannot tell what his ultimate status will be, it is best to omit that man's record from statistical consideration. It is possible to make a fairly adequate correction for experience by studying learning curves in the particular job for a considerable number of workers. If we find that the average person after a month at the job is doing about 80 per cent of his maximum

efficiency or that ultimately he will be about $\frac{1}{4}$ better than he is at that time, we may be justified in taking any worker's rating after a month's experience and increasing it by $\frac{1}{4}$ in order to obtain his criterion score. Although this procedure is open to some question, it probably is better than letting the data go uncorrected. An experienced foreman who has trained many men will sometimes be able to estimate fairly well the ultimate status of a man who is in the earlier stages of learning the job. This practice, however, is not to be recommended; it is much better, if possible, to base statistics only on workers who have "arrived."

It may be desirable occasionally to use a more detailed rating form that includes a considerable number of characteristics and lump these together into a combined criterion rather than to take a single overall rating of "efficiency on the job." Since this procedure is not widely used, only one project of this sort will be mentioned; it dealt with retail salespeople [5, 54]. This rating blank involved the following characteristics: accuracy, speed, knowledge of merchandise, display of merchandise, sales talk, adjustment to selling, adjustment to customer, satisfying customer, manner, and appearance. Each one was rated by the graphic scale technique. For instance the item "manner" was arranged as follows:

Courteous but indifferent at times

Condescending in manner, particularly toward some customers

Manner is definite asset to selling

Interest is to please the customer

This procedure may be desirable in a job as complicated as selling, but in most production jobs the proficiency is usually not broken down into so many distinct items. There is a further problem of combining these different items into a single score, but this will not be discussed here.

Reliability of Estimates. Just as the reliability of objective criteria should be ascertained, it is even more important to investigate the reliability of foremen's estimates. For instance, the foreman may be asked to make his ratings and then at a later time, perhaps in a week or two, when he has partially forgotten the exact details of his original ratings, be requested to go through the process again. If his later ratings correlate well with his earlier—i.e., if the same workmen are rated high in both instances,

and vice versa—his rating is more reliable than if this correlation is small. Furthermore, the ratings made by one foreman may be correlated with those made by another. If they agree closely this indicates high reliability, but if the foremen disagree the reliability is low. In the latter case it is sometimes possible in conference to discover the reason for the discrepancies, such as personal prejudice or overemphasis on some minor aspect of the workman's performance. It may be that one foreman is stressing speed and another accuracy, or that one is rating a man low because he is frequently late or because he is ugly. If these matters can be brought out in conference, it may be possible to revise the ratings somewhat and thus secure a truer indication of the actual ability in question. This will increase the reliability of the ratings.

Further investigation may be made by comparing ratings with other criteria such as production. As a usual thing, reliability is found to be highest for tests, next for production, and lowest for foremen's estimates. A correlation of .58 between the ratings of salespeople by two independent judges is comparatively good as such ratings go [5, 53]. The industrial psychologist will encounter situations where the foremen concerned are simply unable to provide reliable ratings. In such cases, if other more reliable criteria are not available, it is useless to undertake the project because the ultimate value of the tests depends on the criterion by which they are evaluated.

MISCELLANEOUS CRITERIA

In addition to the foregoing there are a number of miscellaneous factors that sometimes may serve as criteria. These are not as universal as production or ratings, and some of them are involved in only a limited number of occupations. In some instances they may be sufficiently reliable to supplement or even replace the above criteria.

Quality of Work. One such factor is the quality of work. Whereas in most production records it is quantity that is noted, there are cases in which it is possible to obtain some indication of quality as well. If a considerable amount of the work fails to pass inspection, the percentage that fails may give some indication of the worker's ability. If consumers return goods on account of defective workmanship, this can often be traced to its source.

If a person handles breakable materials, a larger record of breakage serves to indicate a less efficient individual. Some looms carry automatic counters for recording breakages of the yarn. If the occupation is one that involves the possibility of accidents as in the case of a motorman or taxi driver, the number of accidents reported is sometimes taken as the criterion. The amount of material wasted relative to the amount used, as in cutting leather, may be determined by weighing. In these instances the emphasis is on some aspect of the quality of the work done by the employee.

Amount of Preliminary Training. If the concern operates a vestibule school in which prospective workers receive preliminary training under expert supervision before being put on the actual job, it may be possible to obtain some criterion material from the school records. If the grades attained in the school are based on actual measures of skill in the work at various levels of training, these may contribute information regarding the worker's ability in the job. In some types of work the preliminary training continues until the employee has reached a certain level (as judged by his teachers); not until then does he begin actual service or undertake a special kind of job. An aviation pupil is not allowed to solo until he is judged competent by his instructor. In such instances the length of time taken to train the man up to that point may serve as a criterion. If the men have not all had the same opportunities for instruction, naturally this criterion will be unsatisfactory.

Length of Service. The actual length of time the man has been in the employ of the company is of interest. It may be that the more efficient ones remain longer inasmuch as they are successful and contented. It sometimes happens, however, that persons of high ability do not find a simple job sufficiently interesting and hence do not "stick." Therefore, this factor of length of service should not be taken as a sole criterion without careful scrutiny and it should be compared with actual efficiency on the job. It is often illuminating, however, to study records of length of service or turnover with reference to occupational efficiency and with reference to mental tests.

Advancement. With work of an executive nature the criterion is usually more difficult to obtain. Salary is some index of an

executive's ability. If a man is especially good his salary is generally raised in order to keep him, whereas the inefficient man's salary is not raised and he is offered no openings elsewhere. Salary is, however, complicated by other things, such as length of service with the firm or the man's ability to sell himself to the management. Commissions are perhaps a better criterion than salary because they more definitely reflect output. Advancement in the firm is a related factor. In general, the better man is promoted. However, account must be taken of the fact that some jobs are merely a source of supply for certain other jobs, and that being one of several men promoted through such channels is not as indicative of good ability as being promoted through a less usual channel. Again, the responsibility a man is given is some indication of his ability at his work.

There is one other type of criterion that is occasionally used, but which does not so often apply to actual business concerns. This criterion is membership in various organizations which require some particular achievement for admission. Many professional organizations are of this character. An engineer, a scientist, or a professional man who is admitted to the organizations or societies in his field has probably qualified in some way. In certain studies, being listed in *Who's Who* has proved to be of significance.

These miscellaneous factors should not supplant the criteria discussed earlier. The latter are more universal and generally more valuable. In certain situations, however, some of these miscellaneous factors may be useful by way of supplement.

COMBINING HETEROGENEOUS CRITERIA

In obtaining the criterion it is always well to make as many approaches as possible rather than to put all the eggs in one basket. In view of the fact that the tests ultimately developed will be no more effective than the criterion by which they are evaluated, it is important to overlook nothing which may contribute to the reliability of the criterion. Consequently, it is well when possible to obtain production records, estimates from as many superiors as are competent to make them for the workers in question, and any other data that may be available in the particular situation. According to the general principle of aver-

ages, the more figures available regarding a man's ability in the job the more typical will be the average of those figures. In the ordinary industrial situation there are usually two or three superiors who can estimate most of the men in the group. It is highly desirable to have several, inasmuch as one foreman may rate a particular man high or low because of prejudice and this is partially offset in averaging this estimate with those made by the other foremen. Even though every foreman cannot rate every man because of lack of information, if most of the men are rated in common it is possible to make statistical allowances.

Given several sets of estimates made by foremen, a set of production figures, and possibly some other data, the problem then arises of combining these measures into one, because in correlation procedure it is necessary to compare no more than two things at a time—test and job. It is obviously impossible to average directly production records in the form of pieces per hour with estimates in terms of millimeters on a linear scale. Moreover, the linear estimates made by one foreman may not be directly comparable with those made by another. The first may rate all his men very low in ability while the second may be very lenient. A comparatively high figure assigned a man by the strict foreman may be on a par with that assigned to one of the worst men by the lenient foreman. Consequently, it is necessary to consider means for combining these heterogeneous data into a single set of values, one for each workman.

If the original data are in the form of ranks and are complete, i.e., if every worker has been ranked by the executives and ranked in production, the procedure of combining criteria involves merely averaging the ranks assigned each individual. If, for instance, a man is ranked 3 by one foreman and 6 by another, his average rank is 4.5.

Unfortunately, in the practical situation the data are often incomplete. Some men will be unknown to one of the foremen, but it is desirable to include them in the data rather than discard them, in order to have an adequate number of individuals in the final evaluation of the tests. The difficulty is that one foreman assigns ranks perhaps from 1 to 50, while another assigns them from 1 to 43. Here the worst man in the second foreman's data receives 43, while the worst man in the first foreman's data re-

ceives a more severe penalty—50. If a ranking by one foreman is missing, it makes considerable difference whether it is from the series of 43 or 50. It is quite simple, however, to convert ranks into linear scores. For example, if 43 men are ranked by one foreman these ranks may be considered as normally distributed along a scale of 100 points. Likewise the 50 individuals rated by the other foreman may be similarly scattered along a scale of 100 points. Tables that are available make it necessary only to look up the rank and the number of cases in the sample and read the converted score directly [2, 491]. By this procedure each rank for each individual may be converted into a linear score, and whatever of these linear scores are available for a given workman may be averaged. This procedure obviates the difficulties mentioned above.

Combined Quantitative Estimates. When the data consist of estimates on a linear scale or production figures or are in some other quantitative form rather than in the form of rankings in order of merit, a different procedure is necessary for combining the heterogeneous criteria. If the data are complete and consist entirely of the same sort of thing, such as estimates on the same type of linear scale, it is possible to average the figures for each workman with some validity. Even under these circumstances, however, error may be introduced by the fact that different judges use different standards, some being more lenient and some more severe. If the estimates are not complete, there is much opportunity for unfairness. The omission of the estimate of one man by a lenient foreman who has estimated all his fellows will be a distinct penalty inasmuch as his average will have to be based only on the stricter estimates, while his fellows have a lenient estimate to raise their average. Moreover, when dissimilar criteria in entirely different units are used, such as linear distances on a scale and pieces produced per hour, it is impossible to average such data in the original form just as it is impossible to average pounds and kilograms without converting them to a common basis.

One of the best methods for making such criteria comparable is to use *standard scores*, i.e., to convert the figures of a given kind into terms that are relative to all the figures of that kind—as in

converting each estimate made by a given foreman into terms of all the estimates made by that foreman. If the first foreman's estimates average 70 and the second's average 50, an estimate of, say, 40 by the first would indicate a poorer man than an estimate of 40 by the second, because the first is setting a much more lenient standard and the mark of 40 is much lower relative to that standard. Hence one feature of importance in making such estimates comparable is to determine how much each one is above or below the average made by that foreman; i.e., how it compares with the standard he sets.

Another consideration is whether the foreman bunches his estimates together closely, or scatters them over a considerable range. If two foremen give the same average rating of 60, but one places every man between 30 and 90, and the other rates some as low as 10 and others as high as 110, an individual who is rated 30 by the first is doubtless inferior to one rated 30 by the second. Hence it is necessary not only to consider how much a given rating deviates above or below the average rating made by that foreman, but also to consider this deviation relative to the general scatter or variability of that foreman's ratings. This involves computation of the standard deviation. A brief illustration³ is given in the first part of Table 14.

TABLE 14. ILLUSTRATING STANDARD DEVIATION AND STANDARD SCORES

Workman	Rating by First Foreman	Deviation from Average	Deviation Squared	Standard Score	Rating by Second Foreman	Deviation from Average	Deviation Squared	Standard Score	Average Standard Score
Adams.....	30	-30	900	-1.5	10	-50	2500	-1.25	-1.37
Andrews.....	50	-10	100	-.5	50	-10	100	-.25	-.37
Briggs.....	90	30	900	+1.5	30	-30	900	-.75	+.37
Brown.....	60	0	0	0	120	60	3600	+1.50	+.75
Doe.....	70	10	100	+.5	90	30	900	+.75	+.62
Total.....	300	80	2000		300	180	8000		
Average.....	60	16	400		60	36	1600		
Standard deviation			20				40		

³ The examples used throughout are usually oversimplified in the interest of clarity. Tables should not imply that a study of as few as five cases is customary or valuable.

Suppose the five men whose names appear in the first column receive the ratings by a foreman indicated in the second column. Adams is rated 30, Andrews 50, etc. The average rating is 60. Adams's rating of 30 is 30 less than this 60, or his deviation is -30 . These deviations appear in the third column. If we now neglect the signs and average these deviations, we have 16 as the *average deviation*. It is better practice, however, to compute the standard deviation (σ). Instead of the signs being disregarded, the deviations are squared, thus automatically making all signs plus. For instance, Adams's deviation of -30 squared gives 900 (cf. fourth column). These squares are then averaged and the square root taken to give 20 for the *standard deviation*. The right portion of the table gives similar computations for ratings by a second foreman; these have the same average but are more widely scattered—a standard deviation of 40 vs. 20. This measure fits into the mathematical theory of probability better than does the average deviation and actually occurs in the equation for the normal frequency curve (*infra*). Statistical short cuts in computing it are available [1].

Returning now to the significance of a rating of 30 assigned by the first foreman to Adams compared with a rating of 30 by the second (Briggs), we see that while both are 30 units below the average, it is obvious that Adams stands lower in the estimation of the first foreman than does Briggs in the estimation of the second, because the first foreman's ratings do not scatter as much. The best way to express this fact is to take the ratio of the deviation to the standard deviation, i.e., convert 30 to a standard score. The figures in the third column are divided by 20 to give the standard scores in the fifth, and those in the seventh are divided by 40 to give those in the ninth. We may note that the two ratings of 30 mentioned above yield standard scores of -1.5 and $-.75$ respectively, showing the extent to which Adams's rating by the first foreman actually is below Briggs's rating by the second.

Standard scores are directly comparable, if it is assumed that the ratings made by each foreman follow a normal frequency curve. This is the type of curve obtained in most cases where a large number of persons have been measured in some mental or

physical characteristic. The majority of the people score near the average and the farther we depart from the average in either direction the fewer individuals we find. For example, some intelligence test data for several hundred students are plotted in Fig. 1. We group the scores in classes of 10, lay off these classes along the base line, and for each class erect a perpendicular the

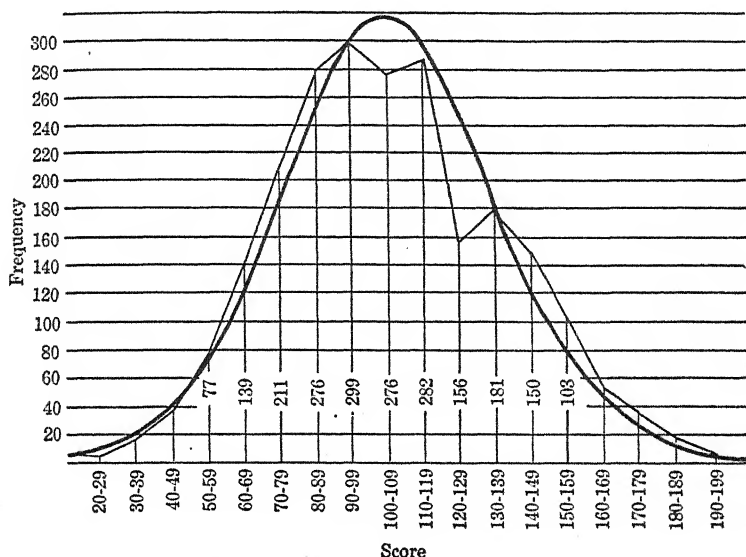


FIG. 1. NORMAL FREQUENCY CURVE

height of which is proportional to the number of students making that score. For instance, 77 students score between 50 and 59 points in the test, 139 students between 60 and 69 points, etc. The line joining the tops of these perpendiculars constitutes the frequency curve. This frequency curve shows the trend above mentioned, with a prevalence of mediocrity and decreasing frequencies as we go toward the extremes. Most curves show minor irregularities like the present one, but the general trend is usually obvious. The ideal normal frequency curve is smooth like the heavy one indicated in the figure. If we know the average and the standard deviation of a set of data we can derive the equation of the curve and plot a smooth one like that shown. Foremen's rat-

ings and production figures usually yield approximately this same type of normal frequency curve.

The equation of the normal frequency curve is known; it is a function of the standard deviation, i.e., the standard deviation occurs in the equation of the curve. The properties of the curve are such that it is possible to tell what proportion of the individuals fall between the average score or rating and any other score, providing this latter is converted into terms of standard deviation. Let us recur to the preceding example of the two foremen each furnishing an average rating of 60, but the first having a standard deviation (σ) of 20 and the second having a standard deviation (σ) of 40. Using these data, we may plot the two normal frequency curves shown in Fig. 2. The ratings are along the base line and the height of the curve at any point represents the proportion of the men receiving the corresponding rating. It is to be noted that both the curves have the same general shape, but the upper one is much steeper than the lower, corresponding to the fact that the first foreman has a smaller variability in his ratings. We now express the scores along the base line in terms of standard deviation (σ) or as standard scores. In the case of the first foreman a rating of 80 represents a deviation of +20, which divided by the σ of 20 gives a standard score of 1.0; i.e., a score of 80 is 1.0 σ above the average.

The equation of the normal frequency curve tells us (by the use of calculus) that between the perpendicular erected at the average and that erected at $+\sigma$ is found 34 per cent of the area of the curve (see figure). This means that 34 per cent of a foreman's ratings fall between these limits. Hence we may say that 34 per cent of the ratings made by the first foreman fall between 60 and 80, and 34 per cent of the second foreman's ratings fall between 60 and 100. Similarly, we know that between a perpendicular erected at the average and one erected at $+2\sigma$ is found 48 per cent of the area of the curve. A workman who is rated 100 by the first foreman is actually exceeded in ability by only 2 per cent of all the men that foreman has rated, and a man rated 140 by the second foreman is likewise exceeded by only 2 per cent of all the men rated. Hence those two workmen have the same ability in the estimation of the two foremen. Tables are available which give areas under the curve for other standard

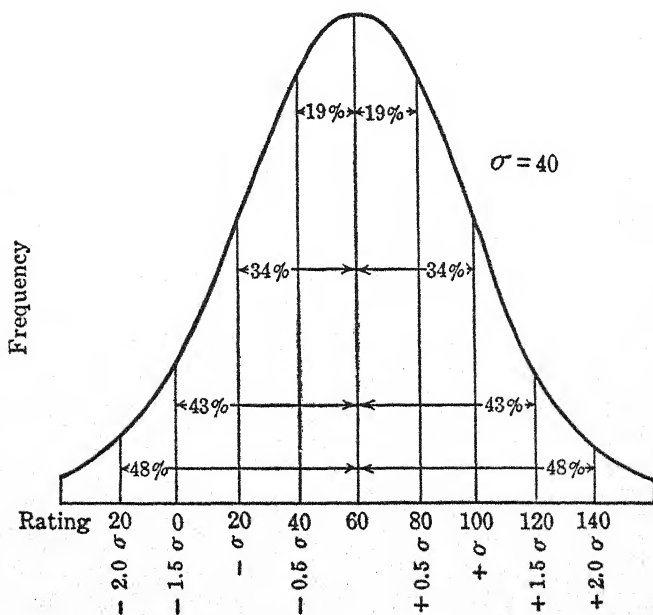
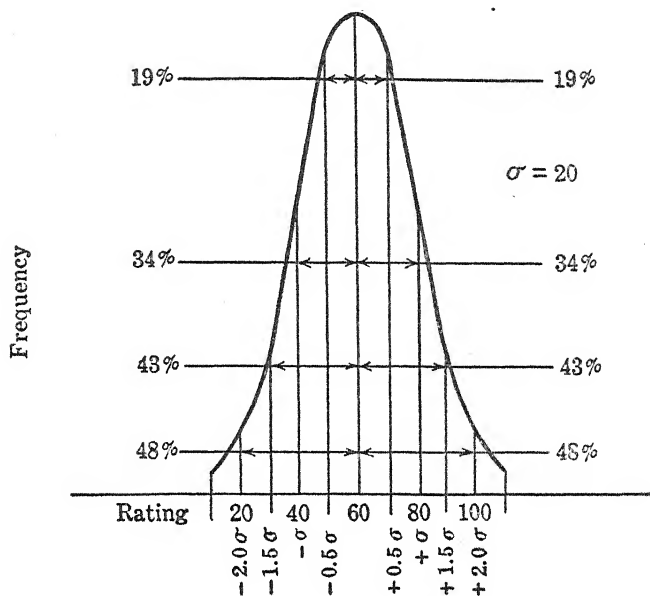


FIG. 2

scores. Hence any such score can be interpreted in a manner similar to that just discussed.

Thus, if we convert original ratings or measures into these standard scores they are directly comparable, because being a certain fraction of the standard deviation above the average in the estimation of one foreman is the same thing as being that same fraction of the standard deviation above the average in the estimation of the other foreman. In other words, we have reduced the measures to common terms, namely, their location on a normal frequency curve, and all normal curves have the same characteristics. Exactly the same procedure may be followed with production figures or any other criterion that can be put in quantitative form. This technique, of course, assumes that the data follow a normal frequency curve. This assumption would be absurd with five cases, as in the above example, which is simplified merely for illustrative purposes. However, if a reasonable number of individuals are involved, the assumption may be made for practical purposes and will make the measures more nearly comparable than if they are treated in an arbitrary fashion. When measures have been converted into this form, it is possible to average algebraically the different measures for a given workman.

In the last column in Table 14 the standard scores in the fifth and ninth columns are averaged.

These techniques make it possible to obtain a miscellaneous set of criteria and combine them into a single measure for each individual. They are all reduced to common terms, and even the omission of some estimates or other criteria in the case of some individuals will not appreciably invalidate the results. The foregoing discussion of combining criterion data by reducing them to standard scores has implied that the different variables are to receive equal weight—that ratings by one foreman are just as important as those by another. It is possible to treat them otherwise. Weights may be based on the judgment of people familiar with the work. If they believe that production figures are twice as important as a foreman's ratings, the standard scores in production can be multiplied by 2. Another possibility is to weight the different criteria according to their reliability, on the general idea that while a particular measure may not be any more important as such, nevertheless the fact that it is more reliable gives it an

advantage and makes it possible to base the decision on something that is a little safer. Whether or not the criterion items are weighted, the combined criterion can be used in validating tests to predict aptitude for a given occupation.

SUMMARY

The criterion is an index of occupational proficiency which is used in evaluating the tests designed to predict that proficiency. It should be derived as carefully as possible because the tests are devised with a specific view to correlation with the criterion and if the criterion itself is inaccurate the entire project is inaccurate. In order to avoid wasted effort, it is advisable to insure at the outset the ultimate availability of adequate criterion data.

Production on the job is the most obvious criterion. After all, production is the factor that it is ultimately desired to predict. In many industrial operations records of actual production per unit time are readily available. Naturally an adequate sample should be obtained—probably an hourly average based on several hundred hours' production. However, several factors may operate to invalidate the production figures unless taken into account. One is the worker's attitude. If he has not done his best at his work, the production is not a true measure of his proficiency. This attitude is assured to a much greater extent in the case of pieceworkers than in the case of day workers, but even here stereotyping of output is encountered occasionally. In evaluating such figures it is necessary to ascertain if the units of production used for different workers are equivalent. If all the men are not making the same product, it is sometimes possible to adjust the figures on the basis of time-study records so that the units can be made equivalent and hence everyone be measured by the same standard. Allowance must be made for extraneous factors, such as temporary technical faults in the machinery and the pace set for the operation by his team-mate if the man under consideration works with another man. Instead of actual production work samples may be employed as a criterion. These are certain standard products made by the worker and evaluated by expert judges. Observation of or acquaintance with the worker is unnecessary. The reliability of production figures or work samples

should be determined by correlating production over one period with that over another or one sample with another respectively.

Estimates as to ability in the job by an employee's superiors constitute a frequently used criterion. This estimate may involve dividing the workers into two groups, but this procedure has little statistical value. It is preferable to select two groups at the extremes of ability, or better still to have a considerable number of groups so that the criterion may more nearly represent a continuous gradation from best to worst. The estimate may also be made by ranking, i.e., arranging the workmen in order from best to worst. This method, however, assumes that the differences between adjacent ranks are equal and often obscures an outstanding instance of superiority or inferiority. A related method is paired comparisons in which each individual is compared with every other one in the group. Estimates on a linear scale are more desirable than the preceding. The name of each workman is followed by a line of uniform length, and the rater makes a check mark at some point along this line to indicate his judgment. The farther from the left the mark is placed, the greater the ability indicated; a measurement of this distance gives the criterion. As a guide to the rater the blank may be divided into columns or have descriptive adjectives or phrases at various positions along the line. The reliability of such estimates should be ascertained by correlating two sets of ratings made by the same foreman on different occasions or by correlating the ratings made by one foreman with those made by another.

In addition to the foregoing there are miscellaneous criteria that are available in some instances. One of these is the quality of work as indicated by the amount passing inspection, breakage, accident claims, or amount of material wasted. Another is the amount of preliminary training given the applicant in a vestibule school or elsewhere before he is put on a regular job or advanced to a particular type of more complex work. Length of service and advancement in the firm give some idea regarding the proficiency of executives and others where it is difficult to obtain other indications of proficiency.

After the various criteria have been obtained, the problem arises of combining them into a single figure for each individual. If the data are in the form of ranks and are complete--i.e., if a

rank is assigned to each man in each criterion—it is a simple process to average the ranks assigned a given individual and obtain his combined rank. If the rank data are incomplete it is necessary to convert them into linear scores (by consulting appropriate tables) before averaging them. If the criteria are in quantitative form, such as estimates on a linear scale or production figures, the best procedure is to convert, for example, estimates by a given foreman into terms of the other estimates made by that same foreman. If the foreman's estimates are averaged, the standard deviation is computed to indicate their variability or scatter, and then a given estimate is converted into terms of its deviation from the average divided by the standard deviation, this estimate is located definitely on a normal frequency curve for that foreman's estimates. It is necessary to assume that his estimates conform approximately to such a normal curve. If estimates of other foremen and production figures are converted into these standard scores, they are all comparable because they are all located on normal frequency curves and the properties of such curves are universal. The measures are thus in common terms and can be validly averaged into a single figure. This combined criterion figure for each workman may then be used in developing tests to predict occupational proficiency.

REFERENCES

1. Garrett, H. E. *Statistics in Psychology and Education*. New York, Longmans, 1937, 493 pp.
2. Hull, C. L. *Aptitude Testing*. New York, World Book, 1928, 535 pp.
3. Laird, D. A. *The Psychology of Selecting Men*. New York, McGraw-Hill, 1927, 345 pp.
4. Paterson, D. G., et al. *Minnesota Mechanical Ability Tests*. Minneapolis, University of Minnesota Press, 1930, 586 pp.
5. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.

Chapter VII

THE SUBJECTS USED IN EVALUATING TESTS

GENERAL CONSIDERATIONS

In devising vocational tests it is necessary to standardize them on a typical group of individuals by comparing test scores with the criterion. The persons used in evaluating the tests are technically called *subjects*. The problem naturally arises as to who shall be used as subjects for the project. Several important considerations are involved in their selection. In the first place, the subjects used in standardizing the tests should be typical of the applicants for employment to whom the tests are ultimately to be given for practical purposes of prediction. Test standards obtained on college students, for instance, would be unsatisfactory for hiring unskilled laborers. Secondly, the incentive or attitude of the subjects should be similar to that involved in the ultimate employment situation. If the test is evaluated on men who do not do their best, the standards will be too low for valid prediction of the capacity of men who exert maximum effort when being tested with reference to employment. In the third place, the previous experience or training of the subjects should be taken into consideration. It is possible that some of the tests will measure factors that are influenced by a man's industrial experience, although they purport to measure innate capacity. In the fourth place, the availability of the criterion must be considered. If it is not forthcoming at the outset, the entire project will be delayed or perhaps vitiated altogether. A further problem arises if a limited group of men of a given type are to be tested, i.e., if some selection of subjects is involved. One must determine how many subjects are necessary and how they are to be selected. Finally, there are several miscellaneous factors to be considered such as the subjects' age, sex, sensory defects, and literacy.

APPLICANTS FOR EMPLOYMENT AS SUBJECTS

Typical of Subsequent Applicants. There are two possible kinds of subjects; applicants for employment may be tested as they are hired for the job in question, or employees who are actually working at that job may be used. The former would seem at first glance to be the logical subjects for the project. They are quite typical of the men to whom the tests are ultimately to be given for employment purposes. The only difference is that the applicants tested at first are an experimental group and their employment does not depend on their test accomplishment, whereas with the later applicants the test may actually determine whether or not they are hired. The applicants in the first group, however, need not know this fact. The general situation in the employment office where the tests are given will be quite similar for both the experimental group and subsequent applicants.

Incentive. In testing applicants the attitude of the subjects is likewise favorable. Inasmuch as they feel that their employment depends to some extent upon their efficiency in the test, they will doubtless have a maximum incentive. This is highly desirable because, as we have previously seen, the only way to keep incentive constant is to keep it maximum. If the experimental applicants and the subsequent applicants both have this maximum incentive their results are directly comparable.

Previous Training. With applicants the uniformity of their previous training with reference to the job in question is apt to be greater than in the case of employees. Sometimes, of course, it is desired to measure actual trade proficiency rather than potential capacity. Usually, however, the psychologist is dealing with occupations for which previous preparation is of little value, for the job involves a new set of specialized operations which the man must be taught, such as building a tire. His interest is not in what training a man has had, but in the innate capacities—attention, motor coordination, or reaction time—that will enable him to be a good tire-builder after he has had requisite instruction at the plant concerned. Whereas in the case of employees some of the abilities that are measured by tests may have been modified somewhat by their work on the job in question, the applicants

are homogeneous in this respect, for they have had no experience of this sort.

Delay of Criterion. The foregoing facts indicate the desirability of using applicants at the employment office as subjects on whom to standardize the tests. There is one serious drawback, however—the criterion will not be available for a considerable time. If the men are tested when hired, it is necessary to wait weeks or months to determine whether they are good or poor in the job. This obviously delays the entire program of evaluating the tests by comparison with the criterion. In the majority of cases this one disadvantage outweighs the advantages of using applicants in the evaluation of occupational tests.

EMPLOYEES AS SUBJECTS

Incentive. Employees are more frequently used in such projects. The question of attitude and incentive is more serious than with applicants, but in the chapter on test technique devices were suggested for controlling this factor. The wording of the test instructions may be such as to impress upon the subjects the importance of doing their best. Their cooperation may be enlisted or they may be effectively motivated by appeals to pride or by competition.

Previous Training. Inasmuch as the employees will have behind them different lengths of service on the job in question, the problem of what the test measures is more acute. Does a particular test measure a man's innate capacity which makes him potentially a good or poor workman at a given job or does it measure traits that have been modified by his training in that job? The general theory of employment tests is based primarily on the first of these alternatives, because applicants come to the office with no experience in the job in question and it is desired to determine whether they have the innate capacity that will enable them to succeed. Hence it is important to determine whether the test measures this sort of thing. The simplest way to obtain this information is to correlate the tests in question with the length of service in the job. If, for instance, the men who have been for a long time employed as tire-builders do better on a certain motor coordination test than do those more recently hired, this indicates

to some extent that the function measured by the test is influenced by experience at that particular job. It is logical then to discard that particular test.

There is another alternative that is comparatively satisfactory. If all the employees used as subjects have had considerable experience in the particular job so that they have definitely reached their maximum efficiency in the job, the presumption is that they have likewise improved in the test function as much as they ever will. Consequently, their results are comparable with one another rather than being invalidated by the fact that one man has had more training by which to profit than has another. Even then, however, this group would score higher than a group of applicants would, and some discount would be necessary in setting a test standard for the latter. Thus it is desirable at the outset to select tests which measure innate capacities rather than proficiencies such as those involved in trade tests. The psychologist is familiar with tests in which the subjects after relatively little practice reach their maximum efficiency. Practice in the test itself would surely improve the function in question as much as practice in the job; and if it can be demonstrated that after a certain amount of practice on the test itself subjects do not improve appreciably, and if that amount of practice is given to those taking the tests, it is reasonable to consider the results satisfactory, regardless of length of service. The latter is then of concern only in evaluating the criterion. At any rate, it is necessary to consider carefully whether the tests measure innate or acquired capacity and to make appropriate adjustments in line with the foregoing suggestions.

Availability of Criterion. With reference to the criterion it is, of course, obvious that this can be obtained almost immediately when dealing with employees. This makes it possible to start at once the procedure of comparing test score and criterion. Such methods as are developed from this procedure can then be put into effect at an early date. If a psychologist is employed to establish such methods of selecting employees, it is probably better to sacrifice slightly the greater reliability obtained by testing applicants in the interest of avoiding delay and getting something of practical value started as soon as possible.

SAMPLING OF SUBJECTS

Number Desirable. Assuming, then, that a group of employees in a given job are to be used for evaluating the tests, there arises the problem of whom to test. The first consideration is how many to include. In occupations which involve a relatively small number of workers, say not over fifty, it is desirable to test them all. If, however, there are several hundred, it is often advisable to take a sampling, i.e., to select some who will be typical and base the results on them rather than to go laboriously through the entire group. From the standpoint of the management the fewer employees taken away from their work the better, provided equally valid results can be obtained. No arbitrary minimum number can be laid down. From the statistical standpoint there is, of course, no danger of getting too many. As the number of subjects increases, the correlation between test and criterion approaches more nearly to the true correlation that would be obtained with an unlimited number. If the correlation is secured with a small group and the procedure repeated with another group, the second result is liable to differ materially from the first. A few anomalous individuals in one or the other group may be sufficient to throw the results out considerably. With larger groups this is less apt to happen because the anomalous cases will be to a greater extent absorbed by the law of averages.

A point exists in a given project, however, at which the addition of further numbers of individuals does not improve matters to any great extent. It is possible to determine this empirically in correlating a specific test score with the criterion by computing the correlation with, say, 50 individuals, then with 60, then with 70, etc., until the addition of 10 more makes little difference in the correlation. Often, however, the psychologist has to state in advance how many men he will need and stick to his statement. In actual practice one occasionally sees reports of research based on as few as 10 individuals. This is probably too small a number to be valuable. Thirty or 40 sometimes prove fairly satisfactory, but it is desirable to have at least 50, preferably more. No definite minimum number can be specified, but it is doubtless better to err in the direction of too many rather than too few subjects.

The situation is a bit different if the small sample is one of

several. Sometimes when tests are being standardized not on employees but on applicants it may be desirable to test, say, the first 20 who are hired, and as soon as the criterion becomes available make a preliminary study of the validity of the tests. This procedure would give some preliminary idea as to the validity of the tests and might suggest minor modifications to be made with the next sample. Another 20 applicants may then be tested similarly and ultimately the tests may be validated with that sample. If the results are essentially the same as with the first sample, the psychologist begins to attach some significance to the trend. If he is careful in interpretation, small successive samples may thus be used on occasion.

Sampling by Foremen. If a sampling of employees is to be made there are various possible methods of selection. The foreman may simply be asked to send over fifty of his men on some convenient schedule. This lets the foreman make the selection, but the value of the sampling is dubious. On the one hand, he may be governed in his selection largely by convenience and send the men who can be most easily spared at a given time—the poorer men. On the other hand, he is liable to go to the opposite extreme and, wishing his department to make a good showing in the tests, send only his best men. Either of these procedures is unsatisfactory because what is desired for statistical purposes is the entire range of ability. There is need for the good, the average, and the poor in order to determine how the different degrees of occupational ability compare with ability in the tests. A correlation coefficient computed when the range of one variable is restricted—as in the present case with only the good workers instead of all degrees of ability—is smaller than it will be if the entire range is included. Unless elaborate formulae are used for correcting this coefficient [2, 225], it is apt to be misleading.

Sampling by the Psychologist. It is far better for the experimenter to make his own sampling. He can then insure that the entire range of occupational ability is represented in the data. If he is making the selection before he gathers the criterion data, it is well for him to secure a list of all the men from whom the selection is to be made and then take them alphabetically, or write their names on cards, shuffle the cards, and pick the desired number at random. This chance procedure will in the long run

insure a "normal" distribution of ability. If the criterion is available before the testing is undertaken, the foregoing procedure may still be followed, or else the selection can be made with a view to getting a normal distribution. The psychologist can select a rather large number of employees near the average score in the criterion, smaller numbers above and below this average group, and still fewer numbers as the extremes are approached. He will have to be governed somewhat by the actual appearance of the data in determining just which ones to select and will have to exercise considerable judgment, but if familiar with normal frequency curves he will have little difficulty in selecting a group whose criteria distribute in normal fashion.

After a sampling has been made, the men on the final list can be examined at times that are most convenient for all concerned. The scheduling of tests will, of course, depend on the local circumstances. In this connection, however, it is well to provide an alternate for each man, or at least an alternate who can be substituted for any one of several in the event of contingencies. A few of the men on the original list may leave before their turn comes or they may be put on the night shift for a few weeks so that it will be necessary instead to test an alternate of approximately the same occupational ability.

MISCELLANEOUS FACTORS

Sex. A few other factors should be taken into consideration with reference to the subjects involved in a research of the above sort. Workers may be of either sex. In the majority of cases all the workers on a given job will be of the same sex and of course no problem arises. If the tests are standardized on men and used for hiring men, sex is not germane. In some instances, however, both men and women are employed in the same job. If there is a sufficient number of each sex it is probably preferable to evaluate them separately from start to finish, just as if there were two separate jobs. From the standpoint of the criterion there is a danger that a foreman will use somewhat different subjective standards in rating subordinates of the two sexes depending on whether he is a misogynist or a philanderer. There is also a possibility that the time-study results will be influenced by

various notions regarding the relative competence or industrial value of the sexes.

From the standpoint of the tests themselves there is likewise the possibility that the scores will be influenced by mental sex differences. Psychologists are at present uncertain regarding these. Experimental evidence shows that sex differences in actual ability of the kind usually measured by mental tests are slight [1]. There are indications, however, that while differences do not exist in the field of ability they do exist in the realm of interest, attitude, and emotion [3]. Most of the experimental data on sex differences are based on pupils in college or school who had nothing particular at stake in taking the tests. It is possible that under the more stimulating conditions of examination in an employment office differences of this more subtle character may influence the results. At any rate, nothing is to be lost by a separate evaluation of the sexes and there is a possibility that error may be thus avoided.

Age. Another factor to be considered is the age of the subjects. If there are certain mental capacities that do not reach their maximum until relatively late in life, or if there are others that begin to decline relatively early, these may make a difference in test results. Consequently, in testing persons in their teens or well along in middle life, it is desirable to consider whether the test under consideration appreciably reflects the age factor.

Many tests have been standardized on people of different ages so that it is possible to plot curves showing how proficiency in the test varies with the subject's age. A few typical results are shown in Fig. 3. The figures along the base line represent age. To make the curves comparable, the average score attained by subjects of a given age is reduced to a percentage of the maximum average score made by any age (in most of the available instances this maximum score is for age 18). These percentages are plotted on the vertical axis. For instance, at age 6 the score in tapping is about 61 per cent of the maximum attained at age 19; at age 7 this percentage has risen to about 66. The heavy straight line indicates the percentage that a given age is of 18 and shows the progress to be expected if development in the various capacities measured is directly proportional to age. The data are taken from various sources involving different groups of persons and

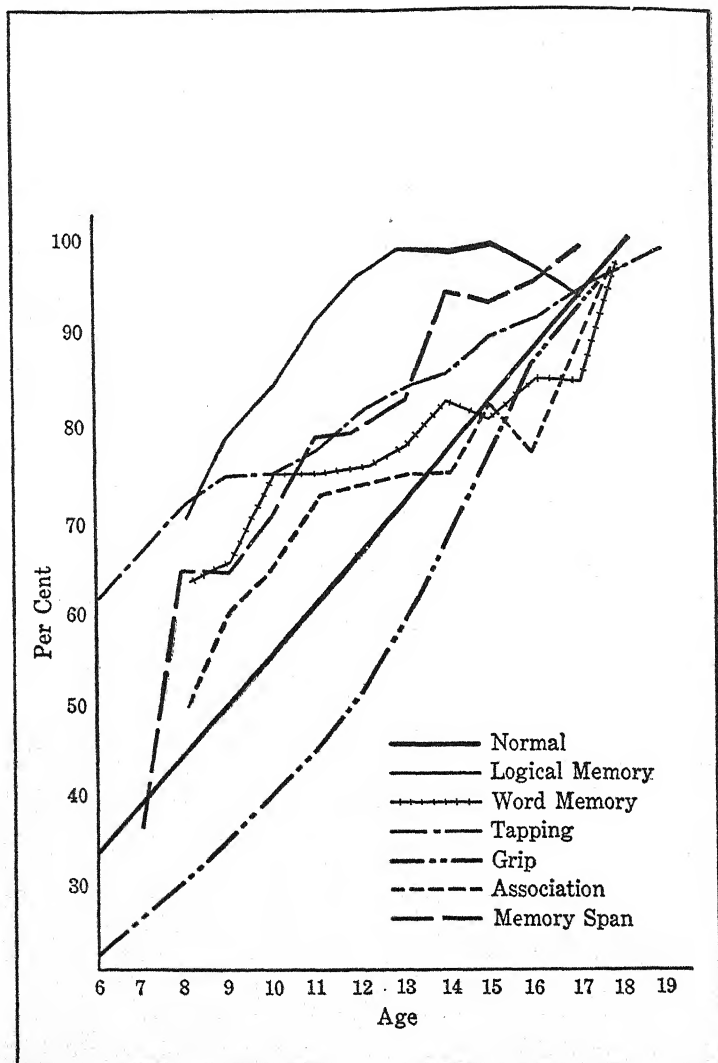


FIG. 3. EFFECT OF AGE ON TEST PERFORMANCE
(Data from Pyle, Gilbert, Smedley, Hollingworth, Whipple)

different numbers in the groups, and include averages for each age based on boys and girls combined. While other experiments and other treatment of data might show somewhat different results, the present curves are probably sufficiently typical to indicate certain trends.

The curves all show obviously a general rise with advancing age. They do not all manifest, however, the same consistent rate of rise. Muscular strength as far as indicated by the grip increases consistently from year to year. Rate of tapping does likewise, although the curve is not as steep, and in the later teens it approaches more closely to maximum proficiency. This suggests that immature workers are more suited to work requiring rapid muscular movement than to work requiring muscular strength. Different kinds of memory show varying rates of increase with age. The simple rote memory measured by memory span rises steadily to its maximum at age 17. Logical memory, however—i.e., memory for ideas in a story that is read—reaches its maximum at about age 13 and remains practically constant thereafter. Memory for disconnected words shows a period of little progress in the early teens followed by a subsequent jump. Consequently, in giving memory tests to young employees or applicants some types of test will presumably be vitiated by the age of the subjects unless allowance is made, while with other types this will not be the case. Free association likewise has a period of little progress followed by a subsequent rise.

These curves are typical of the differences obtained with other kinds of tests and they indicate, with subjects who are not adults, the desirability of taking account of age. It is obviously possible for some applicants to obtain a score which from the standpoint of predicting their ultimate proficiency in the job will be unfair, because at the time of testing their immaturity is conducive to a lower score. In such a project it is desirable to study the scores made by individuals of various ages in a given test and determine at what age improvement due to maturity ceases.

Turning to the other end of the age scale, we find a similar possibility that a person will make a lower score in the test because of senescence. The problem is different from the one just discussed because with the adolescent his score in the test may be inferior to what it would be after he matured, so that

his ultimate efficiency on the job would be predicted too low because the test score itself was too low. If a person at the other end of the scale makes a poorer score in the test it is probable that this "slowing down" would appear in the job itself, so the test score could be interpreted at face value. It would be safer, however, to validate the test with subjects in the middle age range.

A number of studies have been made of the performance of middle-aged and old people on various types of psychological tests. Most of these studies show a moderate decrease in effectiveness in the later years. In dexterity and coordination the loss is reported in some studies as around 10 per cent. The results of one of the more comprehensive investigations are summarized in Table 15 [4]. Three hundred and twenty-four individuals

TABLE 15. PSYCHOLOGICAL TESTS OF VARIOUS AGE GROUPS—MALE¹
BEST PERFORMANCE EQUALS 100 PER CENT.

	Age				
	10-17	18-29	30-49	50-69	70-89
Visual acuity	100	98	96	77	48
Rotary speed	87	100	99	88	77
Reach precision	92	100	98	88	60
Manual reaction	85	95	95	100	70
Foot reaction	76	100	100	95	67
Code memory	75	95	100	64	46
Spatial relations	72	96	100	85	69
"Judgment"	78	100	83	71	56

whose ages ranged from 10 to 89 were given the series of tests that appear in the column at the left. Visual acuity is the ordinary measure obtained with reading charts; the subject was allowed to use his glasses if he was accustomed to them. Rotary speed was the rapidity with which one could turn a small wheel like that on a hand drill or an egg-beater. Reach precision involved reaching, grasping, and placing an object such as a small

¹ After Miles.

block in a hole. Manual and foot reaction involved lifting the hand or foot from a key which broke an electric contact in response to an auditory stimulus. Code memory was measured by filling in letters for corresponding symbols according to a code. Spatial relations involved comparing items of different shape on one part of the page with similar items elsewhere on the page drawn to a different scale. Judgment was a brief intelligence test.

To make the scores in the various tests comparable, the original data are transformed so that the best score for a given test—that is, the score made by the best age group—is 100 per cent. The other scores are taken as a percentage of this. In cases like reaction time where smaller scores are preferable the inferior scores would normally be larger than 100 but are arbitrarily subtracted from 2 to give a score comparable to those of the other tests. For example, if the fastest group in reaction time has an average of .21 of a second this is called 100 per cent. Another group with an average of .26 would actually be 124 per cent of the first. Subtracting 1.24 from 2 gives .76. We may say roughly that the second group is about 76 per cent as good as the first.

The figures for persons beyond the age of 70 show a distinct decrease in efficiency. In the 50-69 group, however, the changes are not nearly so pronounced, but memory as indicated by this particular test has a distinct drop. At least there is enough of a trend to suggest that psychologists administering tests to workers well along in middle age should be cautious and ascertain if age itself is a variable that affects the test scores. Incidentally, in his report on this study the author brings out with considerable emphasis the fact that there is a great deal of overlapping between these different groups. For example, in the 50-69-year group anywhere from 12 to 65 per cent of the individuals do better than the average made by the 18-49-year individuals. This indicates quite clearly that it would be unjustifiable to use age alone as a basis for predicting occupational success as is sometimes done. To be sure, some organizations have retirement plans whereby retirement becomes automatic at a certain age. This is probably based on the theory that individuals decrease in effectiveness beyond that age and the average person of that age would be unsatisfactory on the job. However, these experiments

show that some people well along in years are superior to some of the younger ones. This throws the whole matter back to a measurement of the individual differences, which is the only sensible procedure that a personnel psychologist would follow. Miles in reporting on this project states that "demonstrated individual ability rather than recorded chronological age will dominate the interest and point of view of the personnel directors of the future."

Sensory Defects. Marked sensory defects will make the tests worthless. If verbal directions are used, a partially deaf person will be at a distinct disadvantage; he may not be able to understand what is wanted at all. In a test administered individually or to a small group, this condition will probably be noted by the examiner and proper adjustments made. Visual defects are likewise serious. If a man holds the paper in different positions with apparent effort to focus his eyes upon it, the fact is obvious. Some men will mention the fact that they left their glasses at home. But defects of a less marked degree may nevertheless have an effect in decreasing a person's speed of reading so that his test score does not reflect his actual ability. In lieu of ocular examination it is well to ask the subject if he has ever had any trouble with his eyes.

Literacy. One final point should be considered regarding the subjects—their literacy. Many of the tests used are verbal in character and require that the subject be able to read. The general status of the subjects can probably be ascertained from the employment department or by a casual survey of application blanks. If the literacy of a subject is so low as to handicap him in taking the ordinary tests, recourse must be had to tests of the performance type or at least to tests that involve only isolated numbers or letters.

SUMMARY

In selecting the subjects on whom to evaluate the tests that are to be used in predicting the potential efficiency of prospective employees, it is possible to use either applicants or present employees. The applicants have the advantage that they are typical of the group on whom the tests ultimately are to be used.

They likewise have maximum incentive in taking the tests, as their job is more or less at stake. Their results are not influenced by previous experience on the job in question. Their outstanding disadvantage, however, is that it is necessary to wait weeks or months until they have demonstrated their ability or inability in the job before the criterion is available. If present employees are used, the criterion is available at once. Incentive is not as effectively insured and special effort must be made to provide it. It is also possible that some of the tests are actually influenced by correlating test scores with length of service; if a high correlation is found, such tests may well be discarded. The most common procedure is to use present employees as subjects on whom to standardize the tests.

In some jobs that do not involve many employees it is well to use the entire group as subjects in developing predictive methods for those jobs. In other cases it is necessary to make a selection or sampling of the group. It is theoretically desirable to have a sufficiently large number so that the addition of others will not appreciably change the results. Occasionally several successive small samples have proved satisfactory where a large initial sample was not feasible. It is advisable to have the sampling done by the experimenter rather than by the foreman. The former may make it purely random or, knowing the criterion in advance, may make it in conformity with a normal frequency curve, i.e., comprising more men of average ability and a decreasing number as the extremes of ability are approached. It is important for the sample to cover the entire range of ability.

Several other factors should be considered in some situations. Workers of the two sexes should preferably be evaluated separately. Test scores of persons past middle life should be interpreted with caution because some aspects of mental efficiency decline with age. Likewise tests given to persons in their teens should be carefully scrutinized because of the demonstrated fact that proficiency in some tests reaches its maximum at as early an age as 13, while with others maximum proficiency does not occur until 18 or 19. Defects of vision or hearing may invalidate the results if they pass unnoticed. Finally, the literacy of the subjects often imposes limitations on the types of test used.

REFERENCES

1. Allen, C. N. Research on Sex Differences. *Psychological Bulletin*, 1935, 32, 343-354.
2. Kelly, T. L. *Statistical Methods*. New York, Macmillan, 1923, 385 pp.
3. Landis, M. H., and Burr, H. E. A Study of Conversations. *Journal of Comparative Psychology*, 1924, 4, 81-89.
4. Miles, W. R. Abilities of Older Men. *Personnel Journal*, 1933, 11, 352-357.

Chapter VIII

SPECIAL CAPACITY TESTS: TOTAL MENTAL SITUATION

Preceding chapters have discussed the methods of devising and administering mental tests, of obtaining the criterion and selecting subjects upon whom to standardize the tests. We now turn to the application of the foregoing principles to the actual employment situation. It will be recalled that in the introductory chapter the fundamental principle was laid down that it is necessary to validate the tests by comparing efficiency in the tests with efficiency in the occupation. This procedure must be followed with every occupation for which tests are desired.

TOTAL SITUATION VS. COMPONENTS

While this general principle applies throughout all such problems, there is a considerable difference between jobs in their mental requirements and the corresponding types of test that prove successful. Recurring to the classification given in Chapter IV, we may subdivide the problem into (1) tests of capacity or aptitude and (2) tests of proficiency or achievement. The former functions are presumably innate, and the latter acquired. In the former the concern is with such inborn capacities as attention, memory, or intelligence in so far as they make a man a potentially good tire-builder, even though he has never seen a roller or core; while in the latter the effort is to measure a man's present ability as a carpenter or his ability in some other trade that he has learned. The first of these problems, that of innate capacity or occupational potentiality, is larger. It may be subdivided on the basis of special capacity, such as attention, memory, reaction time, and general capacity or intelligence.

When developing tests of special mental capacity for predict-

ing vocational aptitude two common methods of approach are followed in selecting the tests. On the one hand the entire mental situation involved in the job may be reproduced in a single test. On the other hand the performance may be subdivided into its mental components, these may be measured separately, and the results then combined into a single score. It will be well to illustrate these two approaches in some detail. Tests for automobile drivers will serve as an example [3].

When reproducing the total mental situation for this type of work an endless belt about 10 feet long and one foot wide is mounted on rollers at each end and driven by a motor and gear-reduction unit. The subject is seated at one end of this belt; the top portion, which is visible, moves toward him. Along the edge of the belt are mounted small telephone poles or trees; and as this belt moves toward him, the subject, looking through a miniature windshield, gets a definite illusion of driving down the road. The dimensions of the road, windshield, etc., are worked out to scale so that the actual stimulus on the retina is similar to what it would be in real driving. The subject can speed up the belt by operating a foot accelerator and give himself the impression of driving faster, for the road actually comes toward him more rapidly. With a steering wheel or similar device he can turn the small unit containing the dummy windshield from side to side and thus simulate driving in one lane or the other.

In the right lane a small $\frac{1}{4}$ -inch steel belt riding on top of the main one and driven by a planetary gear arrangement from the main driving gear carries a number of toy automobiles spaced irregularly. The size of these is also scaled to the rest of the apparatus. The planetary gears make it possible for the subject by operating his accelerator to speed up the large belt without affecting the speed of the smaller belt. Thus he is able to "overtake" the small cars visually. In fact, by steering his windshield into the left lane and accelerating he can pass these cars just as he does in actual traffic conditions. If in this process, however, he hits one of the cars, instead of wrecking the apparatus he makes an electric contact which operates an electromagnet and throws the windshield unit up out of the way so that the car passes harmlessly underneath. An electric counter records this "accident." Similarly the small belt which carries the cars is insulated ex-

cept for regions in the vicinity of the car. Brushes rest upon this belt and the circuit is wired so that if the subject "cuts in" too closely an electric counter operates. A similar situation prevails in the left lane. On another small steel belt likewise driven by a planetary gear arrangement are miniature cars coming toward the subject at irregular intervals. Here again as he accelerates he approaches them more rapidly but they do not actually change their speed. If he hits any of them in the process of passing those in the right lane the windshield flies up and an error is recorded. In this way the subject can have a dozen wrecks a minute without actually doing any damage to the apparatus or himself. It is possible, however, to ascertain if he is the type of person who would have wrecks under those circumstances. In administering the tests the subject is given instructions to attempt to pass safely as many cars as he can in a certain time, such as ten minutes.

This test embodies in one single but complicated performance the total mental situation involved in driving a car. The subject has to judge distances and velocity. He must react quickly, for if his reaction time is slow he is apt to strike one of the other cars. There is opportunity for emotional factors to manifest themselves. Some subjects finding themselves in a difficult situation will remain unperturbed and get out of it reasonably well with perhaps only one electric contact, whereas others will become seriously blocked and make a whole series of mistakes, freeze at the wheel, or perhaps drop it altogether and give up.

Tests have been devised for this same job by the other approach of breaking up the job into its mental components. The judgment of distance may be tested by an apparatus which has two rods; one is stationary and the other slides toward and away from the subject who manipulates it by pulling strings. He is required to adjust the movable rod until it appears the same distance from him as the stationary one. His error may be noted on a scale alongside. To measure his judgment of speed as well as distance there may be a small target or a miniature automobile which comes out from behind a screen, moves across an area where it is visible, and disappears behind a screen but continues at the same speed. The subject is instructed to stop it by pressing a telegraph key when it has reached a designated point behind the

screen. It may move at a constant speed or it may be accelerating. At any rate, he has to remember the speed and make his judgment accordingly. Measurements of reaction time may be made in conventional fashion or by using a traffic light as stimulus. When the subject puts his foot on the accelerator a yellow light comes on. At a variable interval of a few seconds thereafter the red light appears, whereupon the subject moves his foot from the accelerator to the brake pedal. Both pedals operate electrically so it is possible to measure the time elapsing between the presentation of the red light and the pressure on the brake pedal. The length of time before the foot leaves the accelerator pedal can also be measured. The emotional factor may be recorded in various ways on the basis of the fact that emotions produce certain involuntary bodily responses. The subject may have apparatus attached to him which records on a moving tape his breathing, blood pressure, and the tremor of his hand when he is startled by a revolver shot or an electric flashover. While everyone "jumps" at such a stimulus, some persons come back to normal quickly whereas others may breathe irregularly for some time, manifest considerable tremor in the hand, and perhaps have a prolonged rise of blood pressure. The time required to return to normal may be an index of emotional stability.

The problem of combining these separate measures into a single score to correlate with the criterion will be discussed later. The present chapter will be concerned primarily with the first of these approaches, namely, reproducing the total mental situation involved in the job. It will also discuss certain preliminary steps and certain statistical treatments of results that are equally applicable in approaching the problem by way of the mental components of the job. This latter procedure, however, will be presented in detail in Chapter IX.

PRELIMINARY PROCEDURE FOR TEST RESEARCH

Establishment of Rapport with Those in Authority. The foregoing discussions of tests and criteria have been of rather general and theoretical scope. A few words are in order regarding the preliminary steps that may be of importance to a psychologist embarking upon a practical program of personnel research in an industrial concern. It is unwise to enter the office at eight o'clock

some Monday morning, give the stenographer some blanks to mimeograph, and send a rating blank to the foreman of the wood-heeling department requesting him to write in the names of his men, rate them, and return the ratings by noon. The foreman may not appreciate what it is all about and be unwilling to cooperate, and the psychologist may be making a mistake in starting with the wood-heelers or even in studying them at all.

A necessary introductory step in undertaking such a project is to get in proper rapport with those in authority in the concern. While a few of the executives who have been instrumental in authorizing the work may know something about it, the other executives and foremen may be distinctly at a loss to understand what is going on. It is well, then, to meet individually or collectively all those who will be in any way concerned with the project and to discuss methods and plans with them. After psychological work has been well established as part of the personnel department it is a different story and the program should sell itself. But when breaking new ground the psychologist should contrive to meet the executives concerned, as well as any others whose cooperation may be needed, and explain the contemplated program. This explanation may well include something about the nature of psychology and its place in industry. The experimental point of view may be stressed, and also the fact that tests are not devised by inspiration or omniscience and immediately put into the employment office but that they must be tried out on employees whose ability on the job is known. This point leads to the necessity for estimates of occupational ability, and these executives may be shown how the entire project depends on the accuracy of these estimates and hence that they themselves are just as important as the psychologist. Furthermore, it is well to point out even in this initial stage that tests are not infallible, that the most they can do is predict probable success, and that there are bound to be cases where a man who gives every indication, on the basis of the tests, of being a successful workman fails to come up to expectations. It is well to make this point at the outset because business people are inclined to attach undue significance to a dramatic instance in which the tests fail and to give insufficient consideration to those cases in which test score and occupational success coincide. This tendency to note the

striking and neglect the typical instances of relations between variables is one of the outstanding fallacies in popular reasoning and accounts for many of our superstitions and other erroneous beliefs. This discussion with the executives naturally should be conducted in terminology with which they are familiar and some effort should be made to "sell" them the psychologist and his program.

Another step in establishing rapport that sometimes is advisable is to have the executives themselves take some tests. The advisability of this depends largely on whether or not they are familiar with tests. In the earlier days this procedure was necessary. At present, with testing playing a larger part in the educational system, a great many persons have encountered tests somewhere in their own school life so that they know what it is all about. If most of them are test-conscious, no particular steps of this sort need to be taken.

Personal Orientation. In addition to securing rapport, another introductory step is for the psychologist to become orientated himself. He needs to be familiar with the whole plant so that he can talk intelligently in any department and he must know the local terminology. Generally it will be in order to spend some time in just going around and familiarizing himself with everything. One function of this process is to locate problems which are amenable to psychological solution. This may have been done somewhat in previous conferences, but for the psychologist who is just starting in to develop employment methods this personal orientation is quite important. When looking for problems he may discover some jobs which require little special capacity of any sort and probably are not worth careful study. He may find others that look quite specialized but where too few people are available for statistical purposes. Still others may necessitate complicated performance with comparatively large numbers of persons. If in one of these latter jobs labor turnover is high, that may be a good place at which to start work, because there is some presumption that psychological methods may be applied successfully.

This kind of observation, in addition to securing orientation and locating problems, may promote a better attitude on the part of all concerned, for the psychologist is seen around and his face

becomes familiar so that there will be less emotional disturbance later when men come to be tested. Moreover, it is in line with morale to discuss problems with foremen on the spot where the problems arise. A few other things are important in getting this personal orientation. Data on labor turnover may be run down to see where, from the standpoint of the management, the most serious difficulties lie. The methods of keeping the payroll may be studied with a view to the possibility of securing production criteria. The general organization may be observed to find who is responsible for various things, so that when anything is needed either in the line of supplies or executive orders the proper person can be approached.

The psychologist must, of course, adapt himself to the circumstances; procedure that will be successful with one group of men and one type of organization may fail with another. But he must definitely strive to get the members of the organization to see his point of view and to understand what he is driving at so that they will cooperate. He must also be familiar with the plant so that he will not make false starts or mistakes because of ignorance of operations or conditions.

ANALYZING THE JOB WITH A VIEW TO SELECTING TESTS

Observation of Workers. Before it is possible to reproduce the total mental situation involved in the job or to devise tests for its mental components, it is obviously essential to find out what elements are involved in the job from the mental and motor standpoint. There are several lines of approach for making such analysis. The psychologist may obtain a good deal of this information by simply observing workmen at the job. If his training has been adequate, he is able to observe people more closely than is the ordinary individual. In a psychological clinic an examiner must be on the alert for significant involuntary movements and traces of emotional instability. He becomes accustomed to going beyond the mere verbal responses of the patient and to interpreting certain mental aspects in the light of what he does. Laboratory training in experimenting on normal individuals will also help in this respect, for the psychologist must watch his subject as closely as the chemist watches the reactions in the test tube. Hence he will be in a position to note whether the workers

are performing their tasks automatically or with apparent conscious effort, whether they have to attend only to one thing or to distribute their attention to a number of things simultaneously, whether they have to exercise a certain amount of judgment or whether the decisions are made for them, whether they apparently take advantage of any rhythm in the operation, and so on. He will also note various other more objective aspects of the work, such as whether it involves large or small muscle groups, whether the time taken to make the motions is critical, whether the men use near or distant vision, and whether they have to remember numbers, symbols, or facts. The psychologist, in short, by virtue of his training will watch the man rather than the machine and will try to analyze the operation from the standpoint of the worker. This type of observation will contribute materially to the discovery of the factors which it is most essential to include in the mental testing project.

Questioning Workers. Further information may often be obtained from the workers themselves. It is usually worth while, unless they are of a very low intellectual order, to talk over their work with them. It is inadvisable to give them a printed questionnaire regarding their work, but a personal interview will often yield valuable results. In an interview it is possible to adapt procedure to the circumstances and to follow a lead when it arises. For instance, one may ask a workman what he thinks about during his work. It may then develop that certain aspects of the job require a good deal of attention. A particular delicate motion may be made with apparent ease, but the man may comment on the difficulty of hitting the hole, thus indicating that a high degree of coordination is necessary. If the worker is asked what he finds most difficult about his work, he may, for example, suggest that he has to be very quick or else "get left." The value of the information obtained in this way will depend largely on the skill of the interviewer. It calls for tact and patience because workers are liable to be skeptical and hesitate to talk about their work. The interviewer must also be sufficiently familiar with the operation to talk to the workman in his own language.

Questioning Executives or Foremen. The information obtained from the workers should be supplemented by questioning their superiors. These latter naturally give the objective sort of infor-

mation that the psychologist himself gets from observing the workers, but the executives and foremen may have been observing the workers for years and have discovered aspects of the job which might escape the psychologist in his briefer observation. Furthermore, foremen have often worked at the job themselves previously and can see it from both standpoints. It may be feasible to ask the foreman or executive to make out a list of the traits which he considers essential for work of this sort, or to describe in detail the qualities of a good worker. It may sometimes be desirable to go to the foreman with a list of traits or qualities and ask him to indicate the ones that are most important in the present connection. Such a list should not be taken, of course, at its face value, but it may well be used as a starting point for further personal discussion. It is sometimes illuminating to get two foremen together and have one "hire" the other, i.e., stage an employment interview. This may yield information of positive value—or quite the reverse, as in one experience of the author's in which it developed that a machinist's proficiency is judged largely by the type of tool kit that he claims to possess.

A good starting point for interviewing a foreman is to ask what is most frequently the trouble with a man who fails to make good at the particular type of work in question. The reply that he does not put his mind on his work suggests the desirability of using some sort of a test for sustained attention. The statement that he is too slow indicates the possibility of using a test that involves reaction time. The procedure to be followed depends upon the individual foreman. As a rule there will be less difficulty in interviewing him than in interviewing the workers, for he will be on the inside and will understand the nature of the whole project.

Personal Experience. It is often advisable for the psychologist to try the job himself. He may have had some laboratory training in self-observation and thus be able to note subtle aspects of the mental state during the job that he would otherwise overlook. He will see for himself just how difficult it is to coordinate, what initial adjustments of attention are necessary, what judgments are involved, how far estimates of space or time are crucial, and to what extent quickness of reaction is essential. It will be illuminating, anyway, to see the job from the inside. A psychologist

who did an extensive piece of research in connection with methods of selecting taxicab drivers started in by driving a cab himself for a few weeks. Although it may not be desirable to pursue a job until a high degree of skill is obtained, undertaking the initial stages of learning may contribute information of value.

Previous Job Analysis. This preliminary psychological analysis is not exactly the same as the procedure of job analysis which is discussed in Chapter XV. The latter is often more comprehensive than that just described and covers such factors as the following: the exact duties involved; the working hours; the general conditions of work with reference to posture, temperature, and hazards; physical qualifications such as strength, vision, or hearing; education; previous experience; amount of judgment and supervision involved, as well as such things as speed or accuracy.

The ordinary job analysis is made by trained interviewers who go over specific topics with workers or foremen and write up the occupational description on the basis of these interviews. Such specifications, however, often involve many things that are, at least by implication, of a psychological character; and if these specifications are available for an occupation the investigator beginning a project of developing mental tests for that occupation will doubtless find the specifications of considerable value. They may not be entirely adequate, especially if made by a person without psychological training, but they will probably call attention to some facts that the psychologist might overlook in his own preliminary analysis. He will naturally go more directly at the psychological aspects of the job. Other possibilities are afforded by the *Dictionary of Occupations* compiled by the U.S. Employment Service. As its name implies, it gives definitions of thousands of occupations based on job analysis techniques; the psychologist may find in a definition hints as to characteristics of the worker that it may be profitable to test. If the job analysis results are available, the psychologist may well take them as a starting point and then proceed still further with the specifically psychological aspects of the job. When he reaches his final conclusion as to the mental factors involved in the occupational situation, he is ready to develop a test or tests for those factors.

The tests for automobile drivers mentioned at the outset of the

chapter were based on analyses of the job like those just described.

One further illustration may be cited—an earlier study of hand-feed dial-machine operators [5, 112]. It will be discussed in enough detail to illustrate the various steps involved in the developmental procedure. These hand-feed dial machines have a series of holes in a rotating table. These holes must be kept filled with material that is to be stamped. The operator has a supply of material and, as the empty holes pass by the point nearest to him, he inserts the material in the proper holes. Analysis of this operation by the procedures mentioned above indicated that it seemed to involve rather sustained attention toward a particular point on the dial and the adjacent portions in the direction from which the empty holes came. It also involved rather close coordination of eye and hand in hitting the hole accurately. Moreover, there was a sort of bodily rhythm in feeding the machine, for it was driven automatically and the holes passed at a constant rate. In this case it proved feasible to devise a single test which reproduced this whole mental situation.

REPRODUCTION OF THE TOTAL MENTAL SITUATION

Simplicity vs. Complexity. After analysis such as the foregoing, the next step is to reproduce the whole situation as far as possible. This necessitates some device that will get the subject into about the same mental attitude that he would have in the actual job. In developing apparatus for such purposes, a number of things must be borne in mind. The apparatus should not be complicated needlessly. This does no harm, but it is unnecessary and involves useless expense and effort. Creditable research has been done with devices constructed of tacks, thread, and pieces of cardboard. Frequently a rather simple device can be made which will give exactly the same effect as a much more complicated machine would, as far as the mental state of the subject is concerned. This point is particularly pertinent because the apparatus at first is purely experimental and may be scrapped if it fails to give results which correlate with the criterion.

Adaptation of Existing Apparatus. Another somewhat similar point is that it is often possible to use or adapt existing apparatus rather than to develop something entirely new. The psychologist

has opportunity to exercise considerable ingenuity in adapting such things as an old phonograph or typewriter to his purposes.

In the above study of the hand-feed dial-machine operators, a large metal disk was mounted on the chassis of a phonograph that drove it at constant speed. Near the margin of this disk were two slots of regulable size. Beneath one point where these slots passed was a funnel. The subject was provided with steel balls which he dropped through the slot into the funnel where they were recorded by a mechanical counter. It was possible to vary the speed of revolution or the size of the opening. Balls which did not go through the slot rolled to another opening where they were recorded separately. The whole device was relatively simple and utilized an existing piece of apparatus rather than building up an entirely new mechanism for supporting the rotating slot. This was highly desirable, particularly at first, because there was no assurance that the apparatus would be permanently used.

Subjective vs. Objective Similarity. It should not be assumed from the foregoing that in reproducing the total mental situation it is necessary actually to reproduce the job on a miniature scale. It is the subjective rather than the objective similarity between test and job that is important. For instance, in a test for street-car motormen it was not necessary to provide a toy car and toy pedestrians. Red and black numbers appearing in different positions at a window in the apparatus were used to produce the rapidly changing mental situation involved in driving the car; the mental aspect was quite similar to that involved in actual practice, although objectively the materials were dissimilar. On some occasions, however, the miniature type of test is preferable for purposes of promotion or in the interest of rapport when this is difficult to obtain.

Fool-proof Character. In devising such a test, the principles discussed in Chapter V should of course be observed. A test for total mental situation generally involves some form of apparatus rather than a printed blank and is perforce an individual test. It is essential with this kind of test to insure its fool-proof character technically. There should be no way in which the subject can beat the apparatus through either cleverness or stupidity. If, for instance, a motion is to be made in only one direction, this can be insured by using a ratchet arrangement so that the

other direction is actually impossible. In a continuous choice reaction test the two telegraph keys can have a rocker underneath so that both cannot be pressed simultaneously. In the test for dial-machine operators the size of the slot was such that two balls could not be inserted at once; any that failed to go through into the funnel rolled to one side and were caught by an apron. Such points as these must be carefully observed in order to avoid a subject's making a higher score than he ought, or a totally unreliable score, by some means which circumvents the experimental situation.

Objective Scoring. In this type of test the method of scoring should receive special consideration. It is undesirable to have the performance one which the examiner must judge qualitatively by observing the subject; it should rather be one that yields some quantitative measure of proficiency. This will usually be in the form of quantity of work done per unit time or time taken for a certain amount of work. In the above example a mechanical counter in the neck of the funnel recorded the number of balls successfully dropped into it; another counter on the disk recorded the number of revolutions or the maximum number of balls that could have been dropped through it. It was then possible to note the actual percentage of efficiency in directly quantitative form. This was far more satisfactory than it would have been to eliminate the counters and determine by watching the subject whether his performance was good or otherwise.

GIVING THE TEST TO WORKERS WHOSE CRITERION IS AVAILABLE

Test Laboratory. It is almost universal practice to install a small laboratory at some place convenient to the plant. In the early days, much testing was done as near as possible to the place where the individual worked, either in a foreman's room or in a portable laboratory that was carried around and set up in the plant. The idea was that taking the worker some distance to a strange place would create an unfavorable attitude and even frighten him. Some investigators went so far as to have a forelady chaperone when they were testing women workers. The difficulties due to the attitude of the subject have now largely disappeared. The younger employees at least have encountered

psychological tests in school and are beginning to accept mental examination in somewhat the same way as they do physical examination. Furthermore, it is common practice to use a shock-absorber test preceding the regular one. This is a test which is not necessarily scored but which removes the novelty of the situation.

A separate laboratory makes it possible to have more uniform conditions in the way of noise, illumination, and ventilation. It is possible to set up the apparatus permanently instead of having to assemble it every morning. Greater flexibility is possible in the tests which are to be employed. A room large enough for a dozen tables for small group examinations and with space for permanent installations of apparatus is desirable. The clerical workers handling the test scores can be in another room if necessary.

Mention should be made again at this point of the desirability of determining a test's reliability (cf. p. 164). If this has not already been established, it is desirable in the present situation either to give the test twice (in a different form if necessary) to each subject or else to provide for separate evaluation of different parts of the test. It is then possible to note whether those who make a high score in one part do likewise in the other, and thus ascertain whether the test is reliable.

CORRELATION OF TEST SCORE WITH CRITERION

After the tests have been given to a group of subjects and scored, the next step is to compare the test scores with the criterion in order to determine whether those who are efficient in the tests are efficient in the job, and vice versa. This makes it possible to state whether the tests are valid and can be used subsequently with applicants to predict their occupational efficiency. The correlation procedure to be described in connection with tests for total situation is equally applicable to the tests for components to be discussed later.

Various methods are available for indicating this correspondence between test scores and criterion. The method that is to be used will be determined somewhat by the form in which the criterion is obtained. If it is possible merely to have the workers grouped into two classes, good and poor, or two classes at the

extremes of ability, about all that can be done is to compute the average test score made by each group. If the good workers make appreciably higher scores on the average than do the poor workers, this indicates something, but the result is not in a form that will enable one to make a very definite prediction of occupational efficiency. If an applicant is given the test, about the only statement that can be made is that his test score is a certain amount above or below the average made by the good workers; it will be impossible to state how big a chance will be taken in hiring him. What is wanted ultimately is some indication of the *probability* of occupational success on the basis of the test scores. This goal necessitates the computation of correlation coefficients. The above method is so inferior to correlation procedure that it will not be discussed further. Every effort should be made to obtain the criterion in such a form that correlations can be computed.

Rank-difference Method. The technique of correlation has already been mentioned (p. 29). It aims to derive a quantitative expression of the tendency for two variables such as test and job to be related so that those who score high in one are apt to score high in the other, and vice versa. One common method of correlation consists of ranking the individuals with respect to each variable and then noting the differences in rank. We may call the best person in the test 1, the next best person 2; similarly, the best one in the job may be called 1, the next best 2, etc. Then for each individual we note the difference between the two ranks assigned him. If most of these differences are small it shows that he is ranked about the same in both test and job, while if the differences are large this indicates considerable discrepancy between his rankings in test and job. From these differences it is possible by appropriate formulae to compute a coefficient which will indicate the closeness of the relationship. Several examples are worked out by this method in Appendix I (Examples I-V); they illustrate not only the method of computation, but also how the correlation coefficient expresses quantitatively the closeness of the relation. When there is a perfect relation the coefficient is 1.00; if there is no relation it is 0. It may even take on negative values as large as -1.00, indicating that the better a person is in one respect the worse he is in the other.

Products-moments Method. The rank-difference method has the drawback mentioned in the earlier discussion of ranks—it assumes that the first person is just as superior to the second as the second is to the third. This often obscures extreme tendencies that ought to be considered. The correlation procedure devised to meet this contingency is called the *products-moments* (i.e., products of deviations) method. It determines essentially whether deviations from the average in one variable are accompanied by corresponding deviations in the other, i.e., whether a person is about as far above the average in one respect as he is in the other, and vice versa. It is necessary to compute these deviations from the average for each individual measure, to get the products of each pair of deviations, and also to compute the standard deviations (p. 191). An example is worked out by the products-moments method in Appendix I, Example VI.

Scatter Plot. The foregoing methods are often quite tedious when a considerable number of individuals are involved. Moreover, it is sometimes desirable, with a set of test scores which are being correlated with a criterion, to observe whether the discrepancies are primarily due to individuals who are poor in the job but who nevertheless do well in the tests, or vice versa. By graphic analysis of the data it is possible to discover any such anomalous cases. It is possible also with the data in graphic form to compute a products-moments coefficient by short-cut methods.

This graphic method involves the construction of a *scatter plot*. The procedure is illustrated in Fig. 4. The test scores in the original data ranged approximately from 1 to 50. Accordingly this range is divided into 10 classes and the rows of the chart are laid off accordingly. For instance, in the bottom row are to be placed men who score between 1 and 5 in the test; in the next to the bottom row are to be placed men who score between 6 and 10 in the test. Similarly, the criterion scores range from approximately 1 to 100; they are likewise divided into 10 classes and the columns of the chart labeled accordingly. In the first column are to be located men who score between 1 and 10 in the criterion; in the next column those who score between 11 and 20. The choice of exactly 10 classes is not essential. In actual practice anything from 10 to 20 classes proves satisfactory.

For illustrative purposes the men in the original data are denoted by the letters A, B, C, etc. Workman A has a test score of 17, which locates him in the row marked 16-20; his criterion is 36, which places him in the column headed 31-40. Only one compartment of the table is determined by this row and this column,

		Criterion									
		1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Test Score	46-50									C	
	41-45								R		G
	36-40								H J		
	31-35						I		L		
	26-30			F		B		P			
	21-25						K				
	16-20			E	A Q		M				
	11-15				O						
	6-10	D	T	S							
	1-5		N								

FIG. 4. SCATTER PLOT FOR CORRELATING TEST AND CRITERION

and consequently A is written in this compartment. Similarly, B's test score of 28 puts him in the 26-30 row; his criterion of 44 puts him in the 41-50 column; only one compartment is determined by this row and column, and B is written in that compartment. In the same way all the other individuals are plotted. In actual practice letters or names are not entered in the chart, but merely a check mark of some sort. If many entries occur in a given compartment, they are subsequently replaced by a single figure which gives the total number of entries.

A glance at Fig. 4 shows a rather definite tendency for the entries to scatter more or less along a diagonal line—from the lower left to the upper right corner. Those in the lower left corner are poor in both test and criterion, while those in the upper right are good in both respects. This array indicates a high correlation between the two variables. With a scatter plot like this, it is possible by short-cut methods to compute the actual prod-

ucts-moments correlation coefficient. In the present instance this coefficient is .90.

By way of comparison two other scatter plots are given in Fig. 5. The class intervals are not indicated; merely the general trend of the distribution is shown. Each dot represents an individual. The chart at the left involves a negative correlation. It is to be noted that the entries scatter roughly along a diagonal

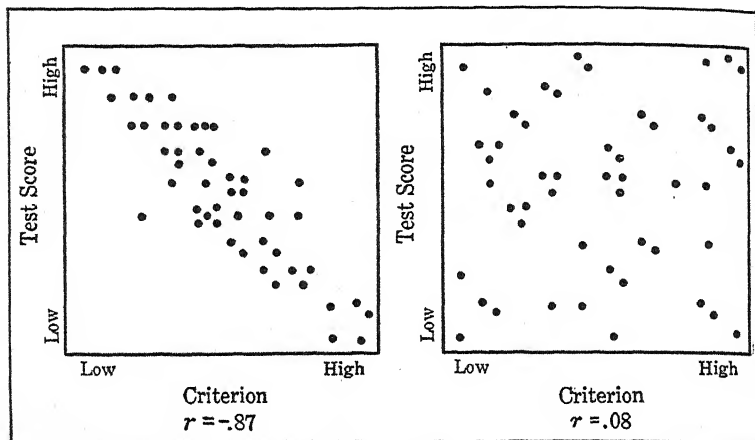


FIG. 5. SCATTER PLOTS FOR A LARGE NEGATIVE AND FOR A SMALL CORRELATION

line from the upper left to the lower right corner. This means that those who are high in test score tend to be low in the criterion, and vice versa. The other chart in Fig. 5 shows the kind of a scatter plot that results from data with a very small correlation. It is obvious that the entries are scattered at random in the plot and there is no tendency for high scores in one variable to go with high (or low) scores in the other.

Methods such as the foregoing afford the best approach to the problem of the relation of test scores to criterion. The particular method used will vary with the nature of the data and the statistical or computing equipment available. The best statistical procedure usually demands a products-moments correlation coefficient in the final evaluation of the two variables.

Whether this is computed by dealing with the actual scores or by plotting them first makes little difference unless one is interested in locating anomalous cases. The main point is to obtain the best possible quantitative expression of the validity of the test, i.e., the tendency to which high test scores go with good ability in the job, and vice versa. In the study of the dial-machine operators mentioned above, the correlation between scores in the test and piecework earnings was approximately .50. This is only a fair correlation, but it indicates some validity for the test.

Regression Equation. It is conventional practice after a correlation has been computed to derive a *regression equation*. This is an equation that expresses criterion in terms of the test and is of the form:

$$X = bY + C$$

where X is the criterion or ability in the job, Y is the test score, and C is a constant term. (Cf. Appendix I, Example VI.) The b term is proportional¹ to the correlation between criterion and test. This equation gives the best estimate that can be made as to ability in the job on the basis of the test. If, for instance, the equation proves to be $X = .6 Y + 20$ and a given applicant for a job scores 80 points in the test, we substitute 80 for Y in the equation, thus: $X = .6 \times 80 + 20$, or $X = 68$. This means that 68 points in the criterion is the best prediction we can make as to his ability.

The solution is, of course, in whatever terms have been used to obtain the criterion. If the latter was obtained in earnings per hour, the equation will predict the most probable earnings per hour, while if it was in terms or ratings on a linear scale, the prediction will be in those terms. The closeness of prediction possible with correlations of different magnitude will be discussed later in the chapter.

¹ The detailed formula is:

$$X - M_x = r \frac{\sigma_x}{\sigma_y} (Y - M_y)$$

where σ_x is the standard deviation of the criterion scores and σ_y the standard deviation of test scores, M_x is the mean or average of the criterion scores and M_y the average of the test scores.

LIMITATIONS OF TOTAL SITUATION METHOD

A test for total mental situation has only one serious limitation as compared with tests for the mental components to be discussed subsequently, and this is the fact that if the test shows a low correlation with the criterion all the work has been wasted. If a rather complicated device has been constructed, workmen have been obtained to take the test, and the final score fails to correlate with ability in the job, it is necessary to begin all over again. This is often rather difficult to do in a particular industrial organization. If the men are called in again for further tests, they naturally wonder why they could not have had these new tests at the outset. It gives the psychologist the appearance of not knowing what he is doing.

In the method of components, on the other hand, a considerable number of tests are given and separately correlated with the criterion. Those that show a low correlation can be scrapped. Some, however, will probably show an appreciable correlation and can be used without the necessity of calling the men back for further examination. The former procedure amounts to keeping all the eggs in one basket, while the latter distributes them so as to minimize the prospect of an utter catastrophe. There are situations, however, in which it is almost certain in advance that a test can be devised which will reproduce the situation and show an appreciable correlation with the criterion. In other cases it may be possible to test the subjects repeatedly without inconvenience. Sometimes the test for total mental situation may be given along with tests for mental components. In all such cases the method is justified.

EXAMPLES

Motormen. Examples have already been cited of a test for the total situation devised for automobile drivers and for hand-feed dial-machine operators. A few other examples may be presented. One is the selection of streetcar motormen in Milwaukee. It involves primarily the ability to react to a complicated series of signals with responses involving combinations of hand and foot such as are necessary on the part of motormen. A signal board comprises seven small openings any one or more of which may

be illuminated by a lamp. The subject has a small stand containing two handles which can be moved backward and forward around a central point, and also two foot pedals. He is given preliminary training in reacting to rather complicated patterns of signals. For example, if light number one is not on but the other six appear, he must pull both handles toward him and step on the left pedal; appearance of all the lights except number four requires a different response of hand and foot [9, 292]. This test was used in hiring motormen in Milwaukee; and whereas in an earlier period 14 per cent of the motormen had been discharged because of accidents, after the installation of the test only 0.6 of one per cent were so discharged.

Punch Press. Apparatus for measuring aptitude for operating a punch press is straightforward in principle [8]. It involves a moving field driven by a motor, and a small punch operated by a foot pedal. Standard pieces of metal are provided with a hole slightly larger than the punch in the center of each sheet. The subject has to feed these through, sometimes from left to right and sometimes vice versa, and have the punch go through each hole without touching the edge. The punch itself is surrounded by a coil spring which permits the punch to stop when the head descends if the stock is not correctly in place or if the subject's finger is accidentally underneath it. A counter records mechanically if the punch meets any resistance, that is, if the subject fails to place the stock so that the punch goes completely through the hole. The test involves about all there is from a psychological standpoint in the actual job. It is largely a matter of judging distances and of timing.

Engine Lathe Aptitude. A test for prospective lathe operators involves two large screws similar to the feed screws on a lathe and mounted at right angles [6]. One of them moves a carriage toward or away from the subject. The other screw is mounted on this carriage so that it moves a second carriage to the right or left relative to the first. The latter carriage has a horizontal extension, the end of which is bent downward to hold a pencil. The pencil writes on a pad at the base of the apparatus. By turning the two feed screws simultaneously, the subject attempts to make the pencil follow a prescribed pathway. The test showed a fair agreement with proficiency in a college course in shop

practice and when combined with some other tests yielded a correlation of .55.

X-ray Machine Operators. One test used in a more professional field may be mentioned, viz., a test for operating X-ray equipment [4]. The mental situation in this job involves following somewhat complicated patterns of switches, essentially a rather difficult directions test. Accordingly this situation was reproduced in a single test with a board which contained a number of point switches. The subject was given complicated directions pertaining to this switchboard, such as: "Turn number one to the right point; turn number two to point three; keep number four closed when the red light is on; when you close number five notice whether you hear a buzz or a bell; if the latter, open the switch immediately."

The foregoing will suffice to illustrate a few projects in which a single test was devised to reproduce the whole situation involved in the job. We shall now turn to interpretation of the test results.

CRITICAL SCORES

After tests for total mental situation or for mental components of the job have been devised and given to operatives of known ability and the final correlation of test or weighted sum of tests with the criterion has been determined, the problem arises as to how the tests are to be used for employment purposes.

A mere regression equation will not be of much use to an employment manager. His real problem is whether or not to hire an applicant, and if mental test scores are available for that man they must be interpreted in easily understandable fashion. Although he may know that the test gives a fairly valid prediction of probable success in the job, he wishes to know the probable success of the particular applicant. This involves the idea of critical score, i.e., a score below which a person should not be hired because of lack of promise of success. This score must be based on the probability that people who fall above it will succeed or that those who fall below it will fail. It is thus important to consider the tests from the standpoint of probability, for, as suggested previously, tests seldom predict with absolute certainty. There are bound to be cases in which a person seems

promising on the basis of the test but fails to come up to expectations. These dramatic instances are apt to catch the attention of the management to the exclusion of the cases of correspondence. Hence it is desirable to emphasize this probability aspect of the critical score as soon as it is determined.

It is possible in each individual case to substitute the applicant's test score in the equation and solve for his criterion score. This gives merely his most probable criterion score. If the equation indicated that he would be doing 79 units per hour after he had learned the job we would know from the general level of production on that job whether 79 would be adequate. But we could not be sure that he would do exactly that amount and we would need to know further the probability of his doing, say, 5 units more or 5 units less. Consequently it is better to generalize so that when we get the validity coefficient for a new test we can determine where to set the critical score in order to achieve a certain probability of the applicant's attaining a given level of proficiency. We also need sufficient generality in order to vary our interpretation in the light of the labor market, because in some cases we shall be forced to make a less rigorous selection. A validity of .60 is not equally valuable in all cases. It is desirable at this point to consider certain aspects of probability as related to the problem of critical scores.

Probability of Occupational Success Predicted from a Regression Equation. When dealing with probabilities, no one hits the mark every time. To draw an analogy from another field, if four coins are tossed the most probable result is two heads and two tails, but if they are tossed repeatedly there will sometimes be one head or one tail and occasionally all heads or all tails. In fact, if the coins are tossed 1600 times there will be approximately 100 cases of all heads, 100 cases of all tails, 400 cases of one head and three tails, 400 cases of one tail and three heads, and 600 cases of two heads and two tails. Thus the best guess as to what will occur in any given toss is two heads, but one cannot be absolutely sure of tossing it. However, one would rather bet that the result of any given toss would be two heads than to bet that it would be three or four heads. That is, the actual values that would be obtained if the event were repeated many times would

average around the most probable value without always coinciding with it.

This same principle applies to the probable value of the criterion computed from a regression equation involving mental test scores. Suppose that a large number of men made exactly the same score in the tests and the regression equation indicated a value of \$60 as the most probable wage. If these men were put to work and after they had learned the job their actual earnings were tabulated, they would average about \$60, but some would be a little more and some a little less. While perhaps the majority would receive \$60, some would receive \$65 and about the same number would receive \$55. There would probably be others, fewer in number, receiving \$70 and \$50, and fewer still receiving lower or higher wages than these. In other words, the actual earnings if plotted in the form of a distribution curve would give the normal type of frequency distribution (cf. Fig. 1, p. 193) with the high point at \$60, the most probable value, and with decreasing frequency the more they deviate from \$60 in either direction. Thus a certain error is involved in estimating one variable from other correlated variables—in this case in estimating success in the job from the tests. This is termed the *standard error of estimate* and is computed by the formula $\sigma\sqrt{1-r^2}$, where σ is the standard deviation of the factor we are trying to predict—i.e., ability in the job—and r is the correlation between job and test. It will be seen that the larger the correlation, the smaller the error of estimate. With a high correlation between criterion and test one can hit the mark rather closely in predicting vocational ability on the basis of the test.

It is possible to use these facts when a given worker is tested in order to determine his probable success. Instead of proportion of the group the same figures can be used to express probability of the individual. We can take his job score computed from the regression equation as the most probable value, and make a distribution curve with this as the average and with the standard error of estimate operating as σ of the curve. This is exactly the same procedure as that outlined in Chapter VI. We can plot a normal frequency curve if we know the average and σ of the measures or estimates. Then, if we lay off the base line of this curve in units of σ , we can determine what proportion of the

cases fall between any assigned limits. (Cf. Fig. 2, p. 195, and the accompanying discussion.) The "cases" in this instance are estimates of job score on the basis of the regression equation. Consider the smaller curve in Fig. 2. Suppose that \$60 is the most probable wage computed from the regression equation and that the standard error of estimate is \$20. If this same result was obtained for a great many men, it is probable that with about 34 per cent the wages would actually (after the men had learned the job) prove to be between \$60 and \$80, because between the average and a value greater or less than it by an amount equal to σ fall 34 per cent of the cases. Similarly, in 48 per cent of the cases the actual earnings would fall between \$60 and \$100, and likewise we should expect 34 per cent between \$40 and \$60 and 48 per cent between \$20 and \$60. To put it in another way, if the solution of the equation for a single applicant gives \$60, the chances are 34 out of 100 that his actual earnings will be between \$60 and \$80, and similarly for any other limits we wish to designate.

We can then decide whether to "take a chance" in hiring the man. Suppose that anyone who will ultimately earn less than \$40 is undesirable; the chances in the present case are 16 (i.e., 50-34) out of 100 that the man will be in that class. Now suppose, in another set of tests which have a higher correlation with the criterion, that the most probable salary is likewise \$60 but the standard error of estimate is only \$10. The probability is then 48 out of 100 that the man will actually earn between \$40 and \$60, because \$40 is 2 σ less than the average and there are only 2 chances out of 100 of his being in the undesirable class of those who make less than \$40. Thus, with this higher correlation and smaller standard error of estimate, we are taking only 2 chances out of 100 of getting a poor man, while with the lower correlation and higher standard error of estimate we are taking 16 chances out of 100. This shows the desirability of high correlations on which to base the prediction of probable success in the job.

Tables for Probable Distribution of Occupational Ability on the Basis of Test Scores. The foregoing method of taking the test scores for each individual applicant and then determining the probable distribution of his success in the job is usually too cumbersome for ordinary employment procedure. The same prin-

ciples may be used for a given job in working out a general table which shows for various ranges of test scores the chances of attaining different degrees of proficiency in the job. Suppose we have 1000 workmen and we divide them into 10 groups on the basis of the tests—the best 100, the next best 100, etc., down to the worst 100. Now suppose we also divide them on the basis of their ability in the job into the best 100, the next best 100, etc. We can then take the best 100 in the tests and note how many of them are in the best tenth in the job, how many of them are in the next best tenth, etc., down to how many are in the worst tenth. Then we can take the second 100 in the tests and see how many of them are in the highest tenth in the job, how many in the next highest tenth in the job, etc. While it is possible, if given enough cases, to construct such a table empirically from the actual data, it is likewise possible, if the correlation between test and criterion is known, to work out such a table in general that will hold for predicting any variable on the basis of another provided they have the correlation indicated. This latter procedure is perhaps somewhat better because such a table—worked out, for instance, for a correlation of .60—can be used in any subsequent vocational situation in which test and criterion correlate to the extent of .60.

A few such typical distributions are given in Table 16. They show the probability of occupational success as predicted from test scores when the correlations are .00, .50, .60, .70, .80, and 1.00. The rows in the table, indicated by roman numerals, give the 10 different degrees of ability manifested in test scores; and the columns, indicated by capital letters, give the 10 degrees of occupational ability. For instance, consider the correlation of .70. Suppose the 1000 men are divided into 10 classes on the basis of their test scores. Class I represents the best 100 and Class II the next best 100. Similarly, Class A represents the best 100 in the job and Class B the next best 100 in the job. The table shows that of those in Class I in the test there will probably be 47 in Class A in the job, 22 in Class B in the job, 13 in Class C, and none in Classes J and K. By contrast, with a correlation of .50, of the men in Class I there are only 32 in Class A, 19 in Class B, and several in Classes J and K. Obviously, with a higher correlation there is less chance that those with high test scores

TABLE 16. FOR INTERPRETING CORRELATION COEFFICIENTS OF DIFFERENT MAGNITUDE

I, II, III, etc., indicate successive deciles (tenths) of test scores; A, B, C, etc., indicate successive deciles of vocational ability.

$$r = .00$$

	A	B	C	D	E	F	G	H	J	K
I.....	10	10	10	10	10	10	10	10	10	10
II.....	10	10	10	10	10	10	10	10	10	10
III.....	10	10	10	10	10	10	10	10	10	10
IV.....	10	10	10	10	10	10	10	10	10	10
V.....	10	10	10	10	10	10	10	10	10	10
VI.....	10	10	10	10	10	10	10	10	10	10
VII.....	10	10	10	10	10	10	10	10	10	10
VIII.....	10	10	10	10	10	10	10	10	10	10
IX.....	10	10	10	10	10	10	10	10	10	10
X.....	10	10	10	10	10	10	10	10	10	10

$$r = .50$$

	A	B	C	D	E	F	G	H	J	K
I.....	32	19	14	10	8	6	5	3	2	1
II.....	19	16	14	12	11	9	7	6	4	2
III.....	14	14	13	12	11	10	9	8	6	3
IV.....	10	12	12	12	12	11	10	9	7	5
V.....	8	11	11	12	11	11	11	10	9	6
VI.....	6	9	10	11	11	11	12	11	11	8
VII.....	5	7	9	10	11	12	12	12	12	10
VIII.....	3	6	8	9	10	11	12	13	14	14
IX.....	2	4	6	7	9	11	12	14	16	19
X.....	1	2	3	5	6	8	10	14	19	32

TABLE 16. FOR INTERPRETING CORRELATION COEFFICIENTS OF
DIFFERENT MAGNITUDE (*continued*)

I, II, III, etc., indicate successive deciles (tenths) of test scores; A, B, C, etc., indicate successive deciles of vocational ability.

$$r = .60$$

	A	B	C	D	E	F	G	H	J	K
I.....	39	20	14	10	7	4	3	2	1	0
II.....	20	19	16	13	10	8	6	4	3	1
III.....	14	16	15	13	12	10	8	6	4	2
IV.....	10	13	13	13	12	12	10	8	6	3
V.....	7	10	12	12	13	12	12	10	8	4
VI.....	4	8	10	12	12	13	12	12	10	7
VII.....	3	6	8	10	12	12	13	13	13	10
VIII.....	2	4	6	8	10	12	13	15	16	14
IX.....	1	3	4	6	8	10	13	16	19	20
X.....	0	1	2	3	4	7	10	14	20	39

$$r = .70$$

	A	B	C	D	E	F	G	H	J	K
I.....	47	22	13	8	5	3	1	1	0	0
II.....	22	22	17	13	10	7	5	3	1	0
III.....	13	17	17	15	12	10	7	5	3	1
IV.....	8	13	15	15	14	12	10	7	5	1
V.....	5	10	12	14	14	13	12	10	7	3
VI.....	3	7	10	12	13	14	14	12	10	5
VII.....	1	5	7	10	12	14	15	15	13	8
VIII.....	1	3	5	7	10	12	15	17	17	13
IX.....	0	1	3	5	7	10	13	17	22	22
X.....	0	0	1	1	3	5	8	13	22	47

[illegible]

will do poorly in the job. The extreme cases of correlations of .00 and 1.00 show this in a still more marked fashion.

Prediction of Success of Individual Applicant. Instead of interpreting the table in terms of the number of the group who will have different degrees of ability in the job, we may equally well use it for a given man who falls in any particular tenth in the tests to predict the chances of his falling in any of the 10 classes in the job. This inference from the proportion of a group to the chances of an individual is a common one. If an actuary finds that 30 out of 100 people of your age and status die before they are 60 years of age, the chances are 30 out of 100 that you will die within that time. Similarly, if test and criterion correlate to the extent of .70, any man whose test score falls among the highest 10 per cent of test scores stands 47 chances out of 100 of being in the highest 10 per cent in the job, 22 chances out of 100 of being in the next highest 10 per cent in the job, etc.

Thus, if the correlation between a particular set of tests and the criterion is known, it is possible by this procedure to work out a distribution like those in Table 16. Then, when an applicant is tested, the examiner can note in which class of test score he falls and compute his probability of attaining the various degrees of occupational success. The determination of a critical score thus involves merely the consideration of how big a chance the management wishes to take.

This may be illustrated by recurring to the example of 1000 men distributed as in Table 16. Suppose that the workmen on the job at the present time who are in the lowest 10 per cent—i.e., Class K—on the basis of occupational proficiency are manifestly unsatisfactory and it is desired in future to hire as few as possible of this grade. Suppose the correlation between test and criterion is .70. Referring to the distribution for this correlation, if we hire from 1000 applicants only the 100 best men in the test scores—i.e., if we establish a critical score between Classes I and II—we shall obviously have no one from Class K. The same will be true of those in Class II in the tests, so if a critical score is established between Classes II and III no one in K will be hired. If, however, the line is drawn between Classes III and IV—i.e., if the 300 best men in the tests are hired—one of them will be in the unsatisfactory vocational Class K. If the

line is drawn between IV and V, 2 out of the 400 men, or .5 of 1 per cent, will be unsatisfactory; and if it is drawn between VI and VII, 10 out of the 600, or 1.6 per cent, will be unsatisfactory. Or suppose that both Classes J and K—i.e., the lowest 20 per cent in occupational ability—are to be avoided, the correlation still being .70. If the critical score is established between Classes II and III—that is, if the best 200 men are hired—only one of them will be undesirable, i.e., .5 of 1 per cent; if the line is drawn between IV and V, there will be 11 such out of the 400, i.e., 2.7 per cent; while if it is drawn between VI and VII, 36 out of the 600 will be unsatisfactory, i.e., 6 per cent. In this way it is possible to see just what percentage of those hired who fall above a certain critical score in the test will be unsatisfactory in the job.

Justification of Efforts to Raise the Correlation Between Test and Criterion. If we now carry through this same reasoning with coefficients of different magnitude, we can see how the value of the tests in eliminating unsatisfactory workers depends on the size of the correlation between test and criterion. Take, for instance, the above problem of eliminating all individuals in Classes J and K—the lowest 20 per cent in occupational ability whom we shall call “unsatisfactory” workers. Suppose the labor market is such that we are enabled to hire the best 20 per cent in the tests, i.e., we place the critical score between II and III. If the correlation is .50, by this procedure 4.5 per cent of the workers we hire will be unsatisfactory; if the correlation is .60, we shall get only 2.5 per cent of such workers, i.e., only about half as many; if the correlation is .70, we shall be accepting only .5 of 1 per cent, while if it is .80, we shall get none at all. These figures appear in the second column of Table 17.

TABLE 17. PERCENTAGE OF UNSATISFACTORY WORKERS
(CLASSES J AND K) THAT WILL BE SELECTED IF
CRITICAL TEST SCORE IS DRAWN BELOW
THE CLASS INDICATED

<i>r</i>	II	IV	VI
.50	4.5	7.5	10.7
.60	2.5	5.0	8.2
.70	.5	2.7	6.0
.80	.0	.7	3.5

In other words, if in this particular instance we can devise a test which correlates .60 with the criterion rather than .50, we almost double our ability to eliminate these unsatisfactory workers, while if we can find one with a correlation of .70 we make only one-ninth as many mistakes as with a correlation of .50.

This type of example makes clear the justification of the effort to obtain a test with as high a correlation as possible. Similar implications will be brought out in the next chapter, where the development of a battery of tests is discussed and considerable labor is involved in "weighting" them statistically with a view to increasing the correlation between the sum of the tests and the criterion. The saving will not always be of exactly the magnitude indicated in the present example, since it will depend on where the critical score is drawn and the proportion that it is desired to eliminate. In the above example, if the critical score is drawn between Classes IV and V or between VI and VII, the results are somewhat different. These facts are embodied in the remaining columns of Table 17. The figures in the columns marked IV and VI are obtained in exactly the same manner as described above for those in Column II. In all these cases the higher correlation manifestly eliminates more of the undesirable workers.

The above discussion sounds largely negative in character. To be sure, the problem often is primarily one of eliminating potential failures. In selecting men for the Air Service, for instance, the main concern is to eliminate the potential "washouts" before they receive instruction; the selection of potential "aces" can take place later during the training program. However, we should not lose sight of the fact that the statistical approach just described makes it possible to select those who are potentially superior workers just as well as to eliminate probable failures. If we wish to locate applicants who are going to be outstanding workers, say in the highest 10 per cent of our present personnel, we have merely to consult the tables and note what critical scores will be necessary in order to insure a high probability that the applicant will be of the superior type desired.

A few other procedures for interpreting test scores may be described. One of them reduces everything to standard scores—that is, deviation divided by standard deviation (cf. p. 190)—

and furnishes a table in which we can look up a person's test score in terms of standard score, note the validity of the test, and then read directly the most probable ability in the criterion in terms of standard score [1, 262]. Two lines from this table are given below by way of illustration. Standard scores normally run from about -3 to $+3$. The cases beyond these limits are negligible. To avoid minus signs, $+5$ is added algebraically to each standard score so that the range is from $+2$ to $+8$ instead of -3 to $+3$. The test scores in these terms appear across the top of the table. The entries in the body of the table give the standard scores for the criterion as predicted from the regression equation for the degree of validity indicated at the left of a given row. For example, with a correlation of .60 as shown in the second row, if the subject makes a standard score of 2 in the test his most probable score in the criterion will be 3.2. Similarly if the correlation is .40 and he makes a standard score of 7 in the test, his probable score in the criterion is 5.8.

<i>r</i>	<i>k</i>	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0
.60	.80	3.2	3.5	3.8	4.1	4.4	4.7	5.0	5.3	5.6	5.9	6.2	6.5	6.8
.40	.92	3.8	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0	6.2

The table also provides information as to the probabilities of scoring between certain limits. The figures in the second column under *k* are the standard errors of estimate. These are essentially measures of variability of criterion scores about the most probable one—about 5.8 in the above example. Here the standard error of estimate is .92. Hence we may write the probable score on the job as $5.8 \pm .92$. This tells us that the chances are 68 in 100 that the subject's standard score will be between 4.88 and 6.72. The chances of the score falling within any other designated limits can be determined by considering those limits as a fraction of the standard error of estimate and following procedures discussed earlier. (Cf. p. 238.) The original table gives data for correlations from 0 to 1.00 by steps of .05.

One more procedure of this type is even more general in nature [7]. It is designed to answer this kind of question: If a given percentage of the present employee group is considered successful and in the proposed test a critical score is set so as

to select a given proportion of the applicants, what portion of the applicants selected will be in the class which is called successful in the job? By extending the theoretical considerations involved in the tables already discussed, it is possible to compile tables to give exactly this type of answer. Only a portion of one table will be given in the present connection; here 70 per cent of the present employees are considered satisfactory. The original article includes similar more extensive tables for many other

TABLE 18. 70 PER CENT OF PRESENT EMPLOYEES SATISFACTORY

<i>r</i>	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.10	.77	.76	.75	.74	.73	.73	.72	.72	.71	.71	.70
.20	.83	.81	.79	.78	.77	.76	.75	.74	.73	.71	.71
.30	.88	.86	.84	.82	.80	.78	.77	.75	.74	.72	.71
.40	.93	.91	.88	.85	.83	.81	.79	.77	.75	.73	.72
.50	.96	.94	.91	.89	.87	.84	.82	.80	.77	.74	.72
.60	.98	.97	.95	.92	.90	.87	.85	.82	.79	.75	.73
.70	1.00	.99	.97	.96	.93	.91	.88	.84	.80	.76	.73
.80	1.00	1.00	.99	.98	.97	.94	.91	.87	.82	.77	.73
.90	1.00	1.00	1.00	1.00	.99	.98	.95	.91	.85	.78	.74

conditions. If we have a validity of .60 between test and criterion we look down the first column until we come to .60 and find the answer in that row. If now we propose to select the best 30 per cent of the applicants we go across to the column which is headed .30. The entry in the table for this row and column is .92; this means that with a test of validity .60 and selecting the best 30 per cent of the applicants, 92 per cent of them will be satisfactory in the sense that they are among the better 70 per cent of the present employees.

Graphic Method. In lieu of this consideration of probable

success computed theoretically from the correlation coefficient, simpler graphic methods sometimes are used. If the criterion consists of only a few groups of occupational ability, such as good, average, and poor, it is possible to plot the test scores of individuals in the three groups and see where the line can be drawn with the least possible overlapping of the groups. This procedure is illustrated in Fig. 6. It shows the determination of a critical score for predicting success in agricultural engineering

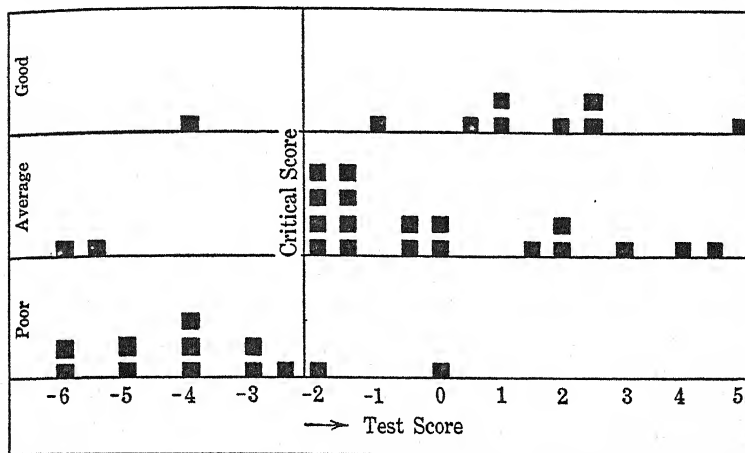


FIG. 6. GRAPHIC DETERMINATION OF CRITICAL SCORE

[2]. The weighted test scores are laid off along the base line and each individual is represented by a square above the appropriate score. The individuals who were considered good by their instructors are plotted in the top section of the chart; those who were rated average, in the middle section; and those who were poor, in the bottom section. After the persons are plotted in this fashion, it is necessary to determine by inspection where to draw a line that will make the sharpest division between those in the poor section and those in the other sections. In the present instance, a line drawn between -2.5 and -2 makes a fairly good division. Only 2 of the poor engineers do better than this critical score, so that there is not a large chance of admitting inferior individuals if such a score is used for vocational advice. On the other hand, only 3 of the average or good engineers fall below

this score, so only a few desirable individuals would be ruled out along with the undesirables.

When the criterion is available in more detailed form and the graphic method is to be used, a scatter plot similar to that in Fig. 4 (p. 231) may be constructed. Recurring to that figure, suppose that practical considerations indicate that workers with a criterion score (salary, foremen's estimates, or what not) of fewer than 41 points are undesirable. We wish to hire as few persons as possible at the left of the vertical line between 31-40 and 41-50 and to hire as many as possible to the right of it. The problem is to draw a horizontal line such that most of those below it will be to the left of the first line, and vice versa. If we use the line between the 16-20 and the 21-25 classes of test score—i.e., if we employ no one who scores less than 21—we shall obviously be eliminating most of the undesirable men. There is only one (F) who falls above this critical score. On the other hand, only one of the desirable men (M) will be eliminated by this procedure. Hence a critical score of 21 points in the test may well be adopted. Persons scoring less than this have little chance of coming up to the requirements of the occupation, while most of those scoring above this amount will qualify.

It should be emphasized again that the determination of critical scores depends to some extent on the relation between the number of applicants for work of a given sort and the number of vacancies. If the situation is such that there are no more applicants than vacancies so that little selection can be made, it is a question of ruling out only the worst prospects and hence a rather low critical score must be used. On the other hand, when the number of applicants far exceeds the number of vacancies so that only a small percentage can be hired, it is to the benefit of all concerned to hire those with the best promise of success. In this case a rather high critical score may be set.

SUMMARY

In embarking upon a program of personnel research in an industrial concern, at least two preliminary steps are desirable. The psychologist must, in the first place, establish rapport with those in authority so that they will be ready to cooperate in every way necessary. To this end the general nature of the

project may be explained to them and they should be shown their own importance therein. It is also well to familiarize them with test procedure by having them take some tests themselves if they have not done so elsewhere. In the second place, the psychologist must orient himself in the organization. He will need to be familiar with the different operations and with the terminology. He may locate departments where there appears to be the greatest need for research and where conditions are favorable for obtaining valid results.

In devising tests of special mental capacity for predicting vocational aptitude, there are two common methods of approach—reproducing the total mental situation involved in the job, or analyzing the operation into its mental components and testing these components separately. In either instance it is necessary to make a preliminary analysis of the mental aspects of the job. To do this, it may be well to observe workers carefully, actually to try the job and observe one's own experiences, to discuss the requisites with foremen and executives with special reference to the distinguishing features of efficient and inefficient workers, or to use as a starting point a job analysis that has previously been systematically conducted. It is also necessary to give the test or tests to workers and to correlate the score or scores with the criterion.

In devising the test for total mental situation, it is wise to avoid undue complexity because at first the apparatus is purely experimental and may later be scrapped. The test need not necessarily be a miniature of the job, for it is the subjective rather than the objective similarity that is important. It should, however, be technically fool-proof and yield an objective score.

The next step is to give the test to subjects whose ability in the job is known. The testing is done preferably in a laboratory set up for the purpose. This affords standard conditions and allows more flexible and permanent equipment. Any emotional attitude toward the tests can usually be controlled by giving a "shock-absorber" test before the crucial series.

After the test has been given to a group of workers, it is necessary to correlate the scores with the criterion. This may be done by appropriate formulae which consider the differences between each subject's rank in the test and rank in the criterion, or in-

volve the product of each man's deviation from the average test score and his deviation from the average criterion score; or the data may be plotted with test scores on one axis and criterion scores on the other. In any instance the magnitude of the correlation coefficient indicates the validity of the test. The next step is to work out a regression equation which expresses criterion in terms of test score and gives the best prediction that can be made of the man's ability on the job with that particular test.

The test for total mental situation has one serious limitation. If its correlation with the criterion proves to be small, the work has been practically wasted and it is necessary to start again. It is often difficult or embarrassing to have the same subjects return later for further examination.

Various examples of such tests were cited. The situation for hand-feed dial-machine operators was reproduced by a rotating disk containing a hole through which steel balls were dropped by the subject. A test for motormen involved a signal board with various lamps and pedals and controls which the subject manipulated in response to complicated patterns of signals. Punch press operators operated a foot pedal to cause a cylinder to go through holes of slightly larger diameter in a metal plate. To measure engine lathe aptitude the subject operated two screws at right angles to make a writing point follow a prescribed pattern. X-ray operators manipulated switches in response to complicated directions. The validity of these and other tests was sufficiently high to warrant their practical use.

After the tests for total mental situation or for mental components have been devised and correlated with the criterion, it is necessary to determine a critical score. This is a score such that persons falling below it will receive unfavorable consideration for employment. The essential thing from the employment standpoint is the probability that the applicant will be a successful worker after adequate training. This may best be determined on the basis of the regression equation. This prediction, however, is not absolute and his actual ability may deviate somewhat from the predicted. But the higher the correlation of test and criterion, the closer will the actual ability come to that predicted from the regression equation. The chances of the

actual ability falling within any particular limits above or below the one predicted can also be computed.

To simplify the interpretation, it is possible to work out for any given correlation a general table showing for various ranges of test scores the chances of attaining various degrees of occupational proficiency. The employment department can then decide where to draw the line for a given set of tests on the basis of how large a chance it wishes to take in hiring applicants, and also on the basis of the labor market. This line is the critical score. Graphic methods may be used for rougher determination of critical scores.

REFERENCES

1. Bingham, W. V. *Aptitudes and Aptitude Testing*. New York, Harper, 1937, 390 pp.
2. Burt, H. E., and Ives, F. Vocational Tests for Agricultural Engineers. *Journal of Applied Psychology*, 1923, 7, 178-187.
3. DeSilva, H. R., and Channel, R. Drivers' Clinic in the Field. *Journal of Applied Psychology*, 1938, 22, 59-70.
4. Falk, E. Eine Apparative Eignungsprüfeinrichtung für Röntgenbetriebe. *Industrielle Psychotechnik*, 1932, 9, 157-160.
5. Link, H. C. *Employment Psychology*. New York, Macmillan, 1919, 440 pp.
6. Patten, E. F. An Experiment in Testing Engine Lathe Aptitude. *Journal of Applied Psychology*, 1923, 7, 16-29.
7. Taylor, H. C., and Russell, J. T. The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection. *Journal of Applied Psychology*, 1939, 23, 565-578.
8. Tiffin, J., and Greenly, R. J. Experiments in the Operation of a Punch-Press. *Journal of Applied Psychology*, 1930, 23, 450-460.
9. Viteles, M. S. *Industrial Psychology*. New York, Norton, 1932, 652 pp.

Chapter IX

SPECIAL CAPACITY TESTS: THE MENTAL COMPONENTS OF THE JOB

As suggested at the outset of the preceding chapter, there are two leading methods of approach to the problem of tests of special mental capacity for predicting vocational aptitude. The first of these—reproducing the total mental situation—was the main topic of Chapter VIII. The other may be termed the method of mental components. The essential feature of this method is the determination of what mental factors are involved in the job and the devising of tests which measure them separately as far as possible. Instead of one test with one final score for the whole mental situation involved in the job, we have a number of tests for the different factors involved and combine them into a single score. Moreover, it is possible to determine the best method of combining them in order to get the most valid prediction.

PRELIMINARY SELECTION OF TESTS

Analysis of Job into Its Mental Components. In order to devise tests for the mental components of the occupation, it is necessary to have some notion of what these components are. This analytic procedure was described in some detail in the preceding chapter. The psychologist may find it profitable to observe the men at work, to talk with them, perhaps to try the work himself, and to discuss with foremen or other supervisors the characteristics of the good and poor workers at this job. If a job analysis has been made, this will often afford valuable insight and give the psychologist a starting point for his own analysis. Even the *Dictionary of Occupations* may be of service. This procedure yields a number of mental factors that presumably are involved in the case of a person working at this occupation.

Devising of Tests Measuring These Components. The next step is to select or devise mental tests which roughly measure these factors. As stated earlier, there is probably no test that measures a single factor, and discrete mental factors may not exist anyway. However, there are occupations in which good attention is obviously a requisite and others which patently require memory, and there are tests which to a considerable extent yield a measure of ability to concentrate and ability to remember, quite apart from the fact that they may measure additional factors. If such tests are selected in the light of preliminary analysis, there is a much greater chance of obtaining some high correlations with the criterion than if tests are selected at random. The number of tests included in this preliminary selection depends largely on the length of time for which each subject will be available. The more tests used, the greater the probability of finding some that are valid, just as the more shots fired at a target, the greater the chance of hitting the bull's-eye. If the analysis indicates relatively few factors that seem obvious, it is well to employ several tests that roughly measure each of these factors, such as several attention tests, several memory tests, or several motor coordination tests, because, while two attention tests may be similar to quite an extent, they may nevertheless vary sufficiently to catch some particular mental aspects that are significant in the job in question.

As an illustration of this procedure we may consider the job of finishing automobile tires [4]. The tire comes to the finisher with several plies of fabric already built on an iron core. He puts it on a frame so that he can spin it by hand, places plies of gum stock on the tread, and rolls them down with hand rollers. In some cases a line is traced around the tire with a pair of dividers and the edge of the stock has to be applied along this line. The workmen said that they had to "keep their mind on the work" in order to be successful. According to the foreman, the men who failed on the job were "too slow." Careful observation of the men at work suggested that they required a rather distributed attention and needed to be able to sustain their attention—i.e., concentrate for a considerable time without a break—and that quick reaction time, good motor coordination, and ability to judge distances were essential. It was feasible to have each employee

for one hour's examination. Consequently, tests that roughly measured the above factors were selected to the extent of one hour's work.

Fifteen tests were chosen for trial. Test 1 consisted of tracing through a series of rectangular patterns between two lines about one-eighth of an inch apart, keeping time with a metronome. Test 2 involved tapping with a metal ring on the tip of the forefinger and making contact alternately with brass plates mounted one above the other two inches apart. Test 3 was similar to Example 2 in Chapter IV. Test 4 was a modification of the one described as Example 1 in Chapter IV. Test 5 involved aiming at a series of crosses on a target with a pencil in time to a metronome; the target was at arm's length and the hand was brought back to the shoulder after every attempt. Test 6 employed a series of shotgun shells made up in different weights to determine the smallest difference in weight that the subject could discriminate. In Test 7 the subject traced a line with a pencil, then drew a line of the same length while the copy and his hand were covered with a screen. Test 8 involved canceling pairs of adjacent numbers whose sum was 10. Test 9 comprised a page of disconnected, unspaced letters—the subject to underline groups of adjacent letters that formed a word. Test 10 involved finding consecutive numbers that were arranged at random (Example 18, Chapter IV). Test 11 was a substitution test, similar to Example 19 (*supra*). Test 12 comprised a series of mazes like Example 21 (*supra*). Test 13 was simple visual reaction time, i.e., the fraction of a second taken on the average to release a key when a stimulus object moved. Test 14 involved watching a moving target which passed in front of an opening in a screen and then continued at the same rate while invisible. The subject was required to stop the invisible target at a designated point by pressing a key. Test 15 comprised a series of lines each accompanied by a short line. The subject determined without measuring how many times the shorter was contained in the longer.

The tests were not actually given in the above order. Those requiring considerable mental effort were interspersed with those that were more motor in character in order to obviate undue fatigue. Each test, moreover, was divided into two installments. The subject went through the first installment of all the tests and

then did the second installments in the same order. This made it possible to compute the reliability of the tests.

FINAL SELECTION OF TESTS

After the tests have been selected, they are given to a group of workers whose occupational ability is known, in the same fashion described above in connection with the test for total situation. The subsequent procedure, however, is somewhat different. In the former case the test yielded a single score and it was simply a question of the extent to which this score correlated with the criterion. In the present case there are many tests and many scores, and it is a question of selecting the best tests and discarding the others. Moreover, some of the tests that are retained correlate more highly with the criterion than do others and consequently should play a larger part in determining the final score. If one test is twice as good as another, scores on that test should be multiplied by 2. This procedure of determining some constant number by which to multiply each score is called *weighting* the tests. It can be shown that if a set of test scores are weighted properly, they will give a better prediction of occupational ability than if they are combined in any other manner. It generally develops that a relatively small number of tests properly weighted will give as good prediction as a large number. Furthermore, the statistical labor in weighting more than ten tests in the best possible way is considerable. It is desirable, then, to select from the large group of original tests a smaller number to be retained for more intensive study.

Preliminary Correlation of Each Test with the Criterion. This selection of the most promising tests is usually made by some preliminary correlation procedure. The method will vary with the circumstances and with the form of the available data. It is not always necessary in this preliminary sorting to employ the relatively laborious products-moments correlation coefficient, for the purpose is merely to eliminate the tests that are absolutely worthless. In some instances a comparison of the average score made by a group of the best workers with that made by a group of the worst will give the desired preliminary information. If the number of workers covered by the study is not too large, the method of rank differences is not specially laborious. With more

TABLE 19. CORRELATION OF PRELIMINARY TESTS WITH ABILITY IN FINISHING TIRES

Test 1.....	.21
Test 2.....	-.09
	.02
*Test 3.....	.31
Test 4.....	-.02
	-.05
Test 5.....	.02
Test 6.....	.01
Test 7.....	-.03
	-.07
*Test 8.....	.49
*Test 9.....	.30
*Test 10.....	.52
*Test 11.....	.35
Test 12.....	.07
*Test 13.....	-.38
Test 14.....	.10
Test 15.....	.06

individuals in the group it is common practice to make scatter plots and determine by inspection which tests are the worst. From the original list of tests the worst ones are eliminated by means of some such methods, and a smaller number, frequently not over ten, of the most promising are retained for further study.

To continue with the example of developing tests for tire finishers, about 50 employees were examined. Estimates of foremen and production figures yielded a criterion score for each workman. The correlations of test scores with the criterion were computed by the method of rank differences. The coefficients are given in Table 19. In instances where two correlations are indicated for a given test, the test was scored by two methods and these were evaluated separately. Obviously, some of the tests are worthless. Consequently the nine tests with low correlations were scrapped and the other six (indicated by stars in the table) retained for further study.

WEIGHTING THE FINAL GROUP OF TESTS

The next step is to determine the proper weight to assign to each of the tests that is retained, i.e., to determine the number

by which to multiply scores in that test before totaling into a single combined score. It might seem logical to weight the tests directly in proportion to their correlation with the criterion. If one test has a validity of .30 and another of .60, the weights might be 1 and 2. This procedure, however, if several tests are to be used, overlooks the fact that the tests overlap one another in varying degrees. Suppose that memory and attention are actually of equal importance in the job, that two tests of memory and one test of attention are retained, and that they all correlate equally with the criterion. If they are all added together with equal weight, twice as much consideration is given to memory as to attention in the final score and employees will be selected preponderatingly on the basis of memory, whereas attention should receive equal consideration. This procedure obviously is unsound for it takes no account of the fact that the two memory tests overlap.

Correlation of Tests with Each Other. This overlapping of the tests can readily be determined by correlating the tests with each other. In the above instance, if scores in the first memory test are correlated with corresponding scores in the second, a high coefficient will doubtless be obtained, while the attention test will not correlate as highly with either of the memory tests. This indicates that the attention test should receive greater weight than either of the others because it is measuring more of a unique factor, whereas the others overlap. When the intercorrelations are known, the next problem is to determine how much allowance must be made for the overlapping. The accepted procedure is the technique of *partial correlation*. Allusion has already been made to this method in connection with the weighting of speed and accuracy (p. 162). For full discussion of the technique the reader is referred to advanced works on statistics [12, 20]. In the present connection effort will be made to present only the general principles and a rudimentary notion of the technique.

Partial Correlation. The scientist is often interested in determining the relation between two things. The chemist studies the relation between the pressure and volume of a gas, the physicist the relation between current and resistance in a circuit, and the psychologist the relation between memory test scores and occupational proficiency. The logical experimental approach to the

problem is to change one of the factors under consideration and note corresponding changes in the other. The chemist varies the volume of the gas to note what happens to the pressure, the physicist alters the resistance in the circuit and measures the corresponding changes in current, and the psychologist selects workmen of varying proficiency in the job and studies their scores in the memory test. However, the scientist must take account of the presence of other factors which may influence the results. He wants to know the actual or intrinsic relation between the factors under consideration quite apart from other things. If the chemist pays no attention to the temperature of the gas, his findings as to change in pressure are as liable to be due to temperature as to volume. If the physicist fails to consider voltage, he does not know whether the change in current is actually due to resistance. If the psychologist takes no account of other factors, such as attention, it is impossible to tell whether the relation between his test and the criterion is due to memory or to something else. The ideal procedure in such cases is to keep the extraneous factors constant. It is possible for the chemist to keep the temperature constant by mechanical means throughout his experiment on the relation between pressure and volume. The physicist can impress a constant voltage on his circuit while he changes the resistance and measures the current.

But there are many problems—and employment psychology faces one of them—in which it is impossible *objectively* to keep the extraneous factors constant. It would be difficult, for instance, to find a group of workmen all of whom have the same powers of attention. In such cases it is possible, however, to control these factors *analytically*. Instead of keeping attention constant by selecting a group of workers with identical capacity, it is possible to test the group that is available and then determine statistically what the relation between the memory test and the criterion would have been if it had been possible to obtain such a select group with constant attention. This involves the derivation of partial correlation coefficients which indicate not the *observed* relation between two variables, but the intrinsic relation between them *with other variables kept constant*.

The ordinary correlation coefficient such as we have been discussing is often quite misleading because of the presence of

other factors besides the two that are correlated. This may well be illustrated by a study made of the relation between hay crop, precipitation, and accumulated temperature [18, 38]. The figures varied when different parts of the year were considered, but the following set illustrates the principles under discussion.

The correlation between crop and precipitation (written r_{cp})¹ was .44, an appreciable correlation—i.e., the more it rained the better the crop grew. The correlation between crop and temperature (r_{ct}), however, proved to be only .05. This did not look right, for common sense says that things grow better in warm weather. Further computation revealed the fact that the correlation between temperature and precipitation (r_{tp}) was $-.44$, i.e., as it became warmer it likewise grew drier. This serves to explain the preceding coefficient of .05. Some relation actually existed between crop and temperature, but this did not appear in the observed data, because when the weather grew warmer, which would naturally tend to increase growth, it also became drier and this tendency worked against the other.

From the above data it was possible to compute a coefficient of partial correlation (the method will be described briefly below) between crop and temperature with precipitation constant. It was obviously impossible to keep precipitation *physically* constant throughout the years when the observations were made. It was possible, however, to control it *analytically* and to determine what the relation between crop and temperature would have been if the precipitation had been kept constant. This correlation ($r_{ct.p}$)² proved to be .30. In other words, there was actually some intrinsic relation between crop and temperature, but it was entirely obscured in the objective data because of the presence of the other factor. When it grew warmer, things tended to grow (as indicated by the partial correlation of .30 between crop and temperature with precipitation constant), but it likewise became drier (as indicated by the correlation of

¹ The common notation in correlation procedure is to write r (the correlation coefficient) with two subscripts indicating the variables correlated—in this case c and p .

² The customary notation with partial correlation is to indicate by the first two subscripts the variables correlated, and by the other subscript or subscripts after the period, the variable or variables kept constant.

-.44 between temperature and precipitation). The net result of these opposed tendencies was no apparent relation between crop and temperature (as indicated by the correlation of .05). This shows how misleading the ordinary type of correlation coefficient sometimes is and how much more illuminating are the partial correlation coefficients.

In scientific study, then, of the relation between two variables, it is desirable to determine their intrinsic relation with other factors as far as possible constant. This principle is especially pertinent in developing a group of tests for the mental components of the occupation. It is desirable to weight each test not in accordance with its ordinary correlation with the criterion, but according to its intrinsic relation as revealed by partial correlation. Suppose, for instance, that three tests are used and the problem is to find the intrinsic relation between the criterion and the third test with the others constant. If it were possible to give the first test to 10,000 subjects, we could find all of those who scored equally in it. Suppose there were 1000 of these individuals. We could give this 1000 the second test and find perhaps 100 of them who had equal ability in this test. With this selected 100 who had identical ability in both of these tests, we could compute the correlation between the criterion and Test 3. We should then have the correlation between the criterion and Test 3 with the other factors constant. It is obviously impossible to adopt such procedure in the employment situation; but it is statistically possible to obtain almost the same result if all three tests are given to the limited group of 100.

The technique of computing partial correlation coefficients is complicated and laborious. A comparatively brief example is worked out in Appendix II. It is necessary to determine not only the correlation of each test with the criterion, but also the correlation of each test with every other test, in order to allow for the overlapping of the tests. From these original correlations it is possible to compute partial correlations like $r_{12.3}$, which indicates the correlation between the criterion (1) and Test 2, with Test 3 kept constant. From this type of coefficient, with one test kept constant, it is possible to compute coefficients with two kept constant, like $r_{12.34}$, which indicates the correlation between the

criterion and Test 2 with both Tests 3 and 4 constant. From these coefficients it is possible to compute those like $r_{12.345}$, in which three tests are constant, and so on according to the number of tests involved.

These computations are all made by formulae like the following:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

where $r_{12.3}$ represents the partial correlation between the criterion and Test 2, with Test 3 constant, r_{12} is the ordinary correlation between the criterion and Test 2, r_{13} is the correlation between the criterion and Test 3, and r_{23} is the correlation between Tests 2 and 3. Suppose $r_{12} = .70$, $r_{13} = .60$, and $r_{23} = .80$. If we substitute in the formula we have:

$$\begin{aligned} r_{12.3} &= \frac{.70 - (.60 \times .80)}{\sqrt{1 - (.60)^2} \sqrt{1 - (.80)^2}} = \frac{.70 - .48}{\sqrt{1 - .36} \sqrt{1 - .64}} \\ &= \frac{.22}{\sqrt{.64} \sqrt{.36}} = \frac{.22}{.8 \times .6} = \frac{.22}{.48} = .46 \end{aligned}$$

In the practical situation interest lies in obtaining the largest possible partial coefficients of test and criterion because they enable a better prediction of occupational ability to be made on the basis of the test. Let us consider what things are conducive to large partial correlations. Suppose that in the above example r_{12} had been .90 instead of .70. The solution of the formula then is:

$$\begin{aligned} r_{12.3} &= \frac{.90 - (.60 \times .80)}{\sqrt{1 - (.60)^2} \sqrt{1 - (.80)^2}} = \frac{.90 - .48}{\sqrt{1 - .36} \sqrt{1 - .64}} \\ &= \frac{.42}{\sqrt{.64} \sqrt{.36}} = \frac{.42}{.8 \times .6} = \frac{.42}{.48} = .87 \end{aligned}$$

The resulting partial correlation of .87 is obviously larger than the .46 obtained in the previous case. This illustrates a fundamental principle, viz., that the larger the ordinary correlation of a test with a criterion, the larger will be its partial correlation with the criterion.

Recurring now to the original example, suppose that r_{12} and

r_{13} had been the same, but that r_{23} had been .30 instead of .80. The solution of the formula is then:

$$\begin{aligned} r_{12.3} &= \frac{.70 - (.60 \times .30)}{\sqrt{1 - (.60)^2} \sqrt{1 - (.30)^2}} = \frac{.70 - .18}{\sqrt{1 - .36} \sqrt{1 - .09}} \\ &= \frac{.52}{\sqrt{.64} \sqrt{.91}} = \frac{.52}{.8 \times .95} = \frac{.52}{.76} = .68 \end{aligned}$$

The resulting partial coefficient of .68 is much larger than the .46 obtained previously and it is due entirely to the fact that r_{23} is smaller. This gives a second principle, viz., that the smaller the correlation of a given test with another test, the larger will be its partial correlation with the criterion.

These two principles indicate what is necessary if tests are to have a high predictive value. If tests which have a high partial correlation with the criterion are desired, those tests are the best whose correlation with the criterion is high and *with the other tests, low*. If two tests show equal correlation with the criterion, but the correlation of the first with the other tests is low, while that of the second with the other tests is high, the first is measuring a more independent factor and its partial correlation coefficient will be higher. It should receive more weight in the final prediction. This, then, is the solution of the problem raised earlier as to how to weight the tests properly in order to obviate the effect of overlapping factors and the danger of giving undue weight to some one factor. The tests are to be weighted not in accordance with their ordinary correlation coefficients with the criterion, but in proportion to their partial coefficients with all the other tests held constant. In this way each test is given a weight according to its intrinsic relation with the criterion. It can be shown statistically that this weighting is more valid than any other that may be devised.

Regression Equation. The actual process of weighting involves the derivation of a regression equation. This is the same equation as was described in the preceding chapter when ability in the job was expressed in terms of score in the test. In the present case, however, the criterion is expressed as a function of several tests. The equation is of the general form:

$$X_1 = b_{12}X_2 + b_{13}X_3 + b_{14}X_4 + \dots + C$$

in which X_1 represents the criterion, X_2 represents the score in Test 2, X_3 represents the score in Test 3, etc.; b_{12} is the weighting for Test 2, b_{13} is the weighting for Test 3, etc.; and C is a constant term. The b terms³ are, roughly speaking, proportional to the partial correlations— b_{12} is proportional to the partial correlation of Test 2 with the criterion when all of the other tests are constant; b_{13} is proportional to the partial correlation of Test 3 with the criterion when all of the other tests are constant. If the equation, for example, proved to be:

$$X_1 = 7 X_2 + 9 X_3 + 14$$

and an applicant scored 12 points in Test 2 and 11 points in Test 3, we would substitute as follows:

$$X_1 = 7 \times 12 + 9 \times 11 + 14 = 197$$

The criterion score of 197 is the best statement of the man's probable efficiency in the job that can be made on the basis of the two tests.

Coefficient of Multiple Correlation. The *coefficient of multiple correlation* is the correlation of the weighted sum of the tests with the criterion. That is, if all the original measures are reconsidered and each is weighted according to the regression equation, these weighted scores can then be correlated with the criterion to obtain the coefficient of multiple correlation. This can also be computed statistically from the partial coefficients without recurring to the original data; it is often computed in both ways as a check on the work.

This coefficient of multiple correlation tells just how valuable the tests are when combined in this manner and makes it possible to see how much superior the combined weighted score is to the score in any single test. This procedure is also useful in determining the minimum number of tests that will give valuable results. If ten tests, weighted, give a multiple correlation with the criterion of .60, and four tests give a correlation of .58, it is probably unwise to retain the entire ten, for four will do nearly as well and much less time will be required to give and

³ These b terms also take into account the variability of the different tests. The C term results from the fact that the equation is first derived in terms of deviations of scores from average score and then transformed into terms of actual test scores.

score them in the employment office. It can also be shown that the coefficient of multiple correlation is higher when the tests are weighted according to the regression equation than when they are weighted in any other manner.

In the previous example of the tire finishers, it will be recalled that 9 of the 15 original tests were discarded on the basis of preliminary correlations. The remaining 6 tests (starred in Table 19) were correlated by the products-moments method with the criterion and with each other. When the regression equation was derived, it developed that the weighted sum of the tests correlated with the criterion to yield a multiple coefficient of .63. This is considerably higher than the validity for the best single test, viz., .52 for Test 10. Further study indicated that almost as satisfactory results could be obtained with only three tests—numbers 8, 10, and 13. Their multiple correlation is .61 and their regression equation is:

$$X_1 = .02 X_8 + .03 X_{10} - .014 X_{13} + 1.82$$

where X_8 is the score in Test 8, etc. The detailed procedure of deriving this equation is given in Appendix II.

Instead of selecting a number of tests and deriving the equation for this entire number, the method can be modified statistically so as to add the tests to the battery one at a time, noting how much the multiple correlation is increased as each one is added. This procedure sometimes is designated the *multiple ratio* technique. In this way we can determine empirically how many tests to include in the equation instead of making an initial guess. An additional advantage of this procedure is that at each stage we are able to correct the multiple correlations for the chance errors added by each test. Usually the increase in the multiple correlation becomes less and less as tests are added, but at the same time the chance error increases [33, 245]. A point is finally reached where the inclusion of another test adds more chance error than actual validity to the battery of tests. Obviously, it is then time to stop. This procedure greatly simplifies the statistical work and it also is a straightforward method of selecting the tests which will go into the battery with the definite objective of having as high a validity as is obtainable with that group of tests.

Another consideration is sometimes important in a battery of tests. A particular factor or capacity which is tested may be absolutely necessary for success in the job, whereas others are desirable but not absolutely essential. This situation normally would be reflected in the correlation coefficients and this absolutely necessary factor would be the first one selected for the battery. However, if it is obvious on empirical or theoretical grounds that this capacity is absolutely essential, the logical thing to do is to ascertain at the outset whether the subject possesses this capacity; if he lacks it, nothing further need be done. For example, a potential operatic tenor might have good breath control, pitch sensitivity, enunciation, phrasing, and so on, but be unable to sing a high note. The initial discovery of this particular point would save a lot of further unnecessary investigation. So in vocational test procedure the psychologist may consider not merely introducing the tests into the equation in a certain order but actually *giving* them in a certain order. This is sometimes called the "successive hurdles" method. If the subject clears the first hurdle he takes the second one, etc. If, however, he falls below a critical score in the first test he is given no further consideration for the job in question. This procedure makes for considerable time-saving in test administration.

Once the optimal weighting for the tests in a battery and the multiple correlation have been determined, the interpretation is exactly the same as in the case of the single total-situation test discussed in the preceding chapter. The regression equation including all the tests enables the prediction of the most probable criterion score. The multiple correlation being known, a critical score for the weighted sum of the tests can be set in terms of the probability of an applicant's achieving any designated level of efficiency in the job.

GENERAL FACTORY OPERATIVES

It is now in point to mention a few projects in which tests have been developed along the lines just discussed. A comprehensive review of the numerous test projects reported in the scientific literature would be prohibitive. However, in order to show the general scope of investigations, a few examples are presented for

general factory operatives, clerical workers, drivers of vehicles, public service employees, salesmen and executives.

A detailed study has been reported of tests for electrical substation operators [42]. Job analysis indicated that the worker had to recall complex series of instructions, to follow directions accurately, and to distribute his attention. The tests employed included some of the McQuarrie mechanical tests, puzzle boxes, tests for following directions in switching operations, a learning test which was a sort of a choice reaction, and a test of persistence. The criterion consisted of ratings and also a record of operating errors. It was possible to set a critical score of 75 points in the battery of tests which would have admitted only 8 per cent of the poorest workers but 71 per cent of the best operators.

A test for garment-machine operators gave particular attention to the criterion [28]. The subjects were girls in trade schools and two types of criteria were used, one depending on quality and the other on speed. The first one utilized work samples such as stitching on tape without a thread, straight stitching on cloth, sewing patches on cambric. Some samples were judged by objective scoring keys and others by qualitative ratings. The keys had reliabilities from .80 to .89 and the experts' ratings from .78 to .83. The speed criterion consisted of the time required for these work samples and had reliabilities between .76 and .88. Nineteen tests were evaluated against these criteria. The highest individual validities were .39 for one test with the quality criterion and .46 with the speed criterion. It was possible to select 6 tests by the multiple ratio technique which yielded a correlation of .57 with the quality criterion. Another battery of 5 tests correlated .64 with the speed criterion. The first battery included matching names (Minnesota Clerical), tracing test, weaving test, paper folding, Minnesota Spatial Relations, and a paper form board. The second included three tests of dexterity, and also tracing and the Minnesota Clerical Test for numbers, i.e., noting whether the numbers in a pair are identical. The interesting contribution of this article is the fact that it suggests validating tests with two criteria and actually setting up two batteries, one to predict speed and the other quality, on the ground that in some industries such as the garment trades there is stress on speed in one plant and on quality in another.

A series of tests were developed for assemblers of electrical fixtures and radios [39]. They included dexterity tests, several visual tests, inserting a stylus through large holes in a plate, and an intelligence test. It was possible to select four tests which, properly weighted, yielded a correlation of .60 with pooled ratings of proficiency on the job.

Tests for garment-machine operators were developed with subnormal girls with I.Q.'s below .70 [40]. After evaluating 17 tests, 5 were retained; these included cutting a design between converging lines, a maze, a paper-folding test, card sorting, and a form of coordination test in which three small targets were tapped in succession repeatedly. A multiple correlation of .66 was obtained and a critical score was set that would eliminate 76 per cent of the failures.

A number of instances are reported where a single standardized test was evaluated for a group of employees and proved quite valid [15]. The Minnesota Paper Form Board Test was used in selecting apprentices for pressmen. Instructors rated them on a 5-point scale and the validity of the test was .58. A standard mechanical assembly test, in which the subject has to put together a number of small objects, was used with cotton-mill machine fixers. The correlation with overseers' rating was .42 for loom fixers and .78 for a small number of spinning-frame fixers [16]. A peg board—that is, a board with holes in it and a series of one-inch pegs to be fitted into these holes with one or both hands—was used in an electrical manufacturing establishment. Here the effort was to predict capacity not in a single occupation but rather in certain broad classes of work such as bench work or coil winding. The correlations with production were around .30 but were consistent [17]. Finger and tweezer dexterity tests in a watch factory showed some relation to salary and length of employment.

One project may be cited to indicate the ingenuity sometimes involved in devising tests. This was a Russian study of the building trades. One item measured was orientation in space and absence of vertigo. The test was taken on a bridge 7 feet above the floor. The subject had to indicate horizontal and vertical directions on papers fastened to the wall while sitting motionless on the bridge or while balancing himself in a swing. Other more

conventional tests were included. An encouraging correspondence was obtained between tests and instructor's ratings of apprentices, particularly in the case of a group of students who had been dismissed by their instructors [43].

A case may be mentioned to show the possibility of utilizing some unusual qualification or characteristic of the worker to the advantage of all concerned. When a subway tunnel was being built, one of the employees who had some "connection" was kept on the payroll although he was lazy and drifted from one job to another. He was, however, of a reckless type and liked to drive a car at high speed. The concrete was sent from the mixer some 500 yards down into the tunnel on a small track. If the truck full of cement went too fast, it tended to jump the track at the curves, and if it was held back by a cable, it tended to stall. Someone hit upon the idea of giving this man the job of riding this truck and controlling it with a brake so as to "make" the curves. This operation appealed strongly to his personality and he was able to deliver the cement at the lower level expeditiously. The output of concrete was increased about 12 per cent and a satisfactory vocational adjustment for the incumbent was achieved [13, 100].

CLERICAL WORKERS

Aptitude measurements for clerical workers have been developed quite extensively. Widely used are the name-checking and number-checking tests in which pairs of names or numbers are given and the subject has to tell whether they are the same or different. Other frequently used items include spelling, detecting errors in a passage, simple computation, indicating where certain proper names belong in an alphabetical list. Other common items are similar to those embodied in abstract intelligence tests. On occasion actual intelligence tests have shown some validity in selecting clerical workers. A tabulation of a number of standard tests is given by Bingham [3, 162]. He presents what are essentially critical scores for different levels of clerical work such as responsible clerical positions, secretarial, and low grade. These scores are given for the Minnesota Clerical Tests, those worked out by Pond at the Scovill Manufacturing Company, one de-

veloped at Carnegie Institute of Technology, a general test used by the U.S. Civil Service, and several intelligence tests.

Two studies deserve a little more detailed discussion. The first of them investigated the validity of a number of clerical and other tests in insurance offices [7]. About 100 workers were involved and the criterion consisted of supervisor's ratings. They covered three grades of work—simple, complicated, and the type that required making decisions. The following tests were evaluated and had the validities indicated: Carnegie Tech Number VI, .41; O'Rourke Senior Classification, .40; Thurstone Clerical, .44; Modified Thurstone, .37; Minnesota Number Checking, .27; Minnesota Name Checking, .29. The rather poor showing for the widely used Minnesota tests is to be noted. The data were also evaluated for a smaller sample with reference to another criterion, namely, promotion after five years or more of service. The validities under these conditions were: Carnegie, .75; O'Rourke, .77; Thurstone, .71; Modified Thurstone, .65; Minnesota Number, .07; and Minnesota Name, .34. The suggestion growing out of this study is that alertness seems to have a little greater validity than strictly clerical tests for this particular kind of work. A further analysis of the subtests led to the general conclusion that those which were distinctly verbal had an average validity of .60; those that were numerical, .52; and those that involved checking, .39. It was suggested that if some of the dead wood in a test were removed, the remainder might have a considerably higher validity. In one instance, by dropping part of a test, the validity was raised from .65 to .79.

A factor analysis was made of a number of clerical and allied tests [1]. The battery included the Minnesota Clerical Test, finger and tweezer dexterity, the Minnesota Placing and Turning Test, a spatial relations test, finding the unique figure in a group of faces, number and letter cancellation, substitution, arithmetic, checking addition. Five factors apparently accounted for everything involved in the intercorrelations. The usual effort was made to speculate as to the nature of the factors in the light of the factor loadings (see p. 107). The first factor appeared to be general ability with some emphasis on clerical aspects. This accounted for about 33 per cent of the total variance in the group of tests. The second factor looked like speed in simple discrim-

ination, for example, the type involved in the cancellation test. The third was evidently spatial, the fourth motor, and the fifth observation and comparison. The first factor, however, carried most of the load. When a similar procedure was carried through with only the more strictly clerical tests such as checking, canceling, substitution, arithmetic, spelling and addition, two factors only were discovered. One was identified as general clerical ability and the other as speed of discrimination.

Brief mention may be made of a few other test projects in the clerical field. In a metal trades organization the Minnesota name- and number-checking tests correlated with a supervisor's rating to the extent of .65 [30]. For the clerical workers in a state reformatory similar correlations averaged around .40 [14]. For bookkeepers, the Thurstone Clerical Test had a validity of .57; when one portion of the test was omitted, it was raised to .74 [34]. A test developed by the National Institute of Industrial Psychology in London was modified somewhat and used with clerical workers in western Massachusetts. It yielded comparatively high validities with supervisor's ratings on rather small samples [25]. Mention should be made of norms on the Minnesota Clerical Test published for a number of occupational groups—viz., accountants, bank tellers, general clerical workers, minor bank officials, shipping clerks [29].

AUTOMOBILE DRIVERS

Tests for automobile drivers were discussed in considerable detail at the beginning of Chapter VIII. Some of these have been used in selecting drivers of commercial vehicles. Some have also been included in an interesting project called a driving clinic. Tests such as visual acuity, depth perception, color vision, coordination, reaction time of various sorts, susceptibility to glare, and judgment of speed were given to subjects who reported voluntarily and to others who were ordered to the clinic by the police. The person's standing in each test in comparison with norms was noted for purposes of individual diagnosis. If the driver had a definite defect such as being considerably slower than the average in reaction time, this was pointed out to him and he was told that unless he made allowance by avoiding situations in which quick reaction was essential he was quite

likely to have an accident. That such a procedure was effective with some subjects was indicated by a reduction in accidents [8].

A successful battery of tests has been reported by an institute for traffic research. The tests involved reaction time, resistance to distraction, general alertness, visual acuity, coordination, and judgment of space, size, and speed. A validity of .77 was found [23]. A similarly high validity of .81 was claimed for some tests developed on 3000 drivers in Barcelona. The criterion was a competitive examination for drivers which presumably included actual tests in driving. The aptitude tests included choice reaction while conflicting stimuli were present, judging velocity of moving rods, and matching numbers with corresponding slots [5]. The results with similar tests for bus drivers in Paris were analyzed for the period from 1921 to 1934. The accidents decreased 37 per cent while the number of vehicles increased 30 per cent. Meanwhile for private drivers where no tests were used, accidents increased 155 per cent and the number of vehicles increased 218 per cent [2].

PUBLIC SERVICE EMPLOYEES

Firemen. Aptitude tests for firemen have been reported with surprisingly high validity (.91), the ratings by the fire chief being used as a criterion. The tests included strength, speed and accuracy of movement, reaction time, vision and hearing, attention, going up a 12-foot ladder across a plank and down another ladder, finding a designated object in a roomful of smoke which was kept "as constant as possible." There were also certain information items which were not strictly aptitude tests [11].

Policemen. Tests [35, 37] have been suggested for policemen with the weights indicated: Army intelligence test (2); accuracy of observation—questions on a picture shown previously and also recording auto tags from memory (1); memory tests—recording facts from a description that is read and identifying photographs shown once and then mixed with other photographs (1); understanding laws and police rules—the subject is given a copy and answers questions by referring to the copy (1); police duties—identifying crimes from definitions and descriptions of cases (1); education and experience (1); personal

traits as determined by an interview (1); medical and physical tests (2).

Postal Employees. A test for mail distributors embodied three parts [27]. The first was essentially an intelligence test. The second was a sorting test which set up a hypothetical arrangement with boxes and names of certain towns and a number in each box. The subject had to give the name of the town to which a number belonged. The third test involved following instructions, centering around a pattern of sorting names. The validity of the combined tests was .71.

Telegraphers. A group of men in a telegraph school were studied with a view to measuring potential telegraphic aptitude [36]. The tests finally used were a rhythm test—writing dots and dashes heard by the subject—conventional opposites, analogies and directions tests, and a completion test—supplying missing words in a paragraph. The multiple correlation for the battery was .53, but the rhythm test alone had a validity of .48.

Telephone Operators. Of somewhat historical interest are some early tests given to girls in a telephone school [26]. They comprised memory span for numbers, i.e., the maximum number of digits that could be repeated after a single reading; a cancellation test—crossing out certain letters on a page; a memory test by the method of word pairs; a test of card sorting; a motor coordination test—tapping rapidly in sequence three crosses on the blank; and speed of association reaction to a stimulus word. The scores of the girls were ranked in each test and the average rank was computed for each girl. These average ranks were then compared with proficiency in the telephone school after three months' service. There was a marked although not universal tendency for those with the better test scores to be more proficient in actual service. The company, moreover, had surreptitiously introduced a number of expert operators among the supposed pupils in the school. These experts made high scores.

Telephone Service. Tests for telephone dial switchmen included some that were essentially trade tests, with questions regarding electrical principles; but there were also tests of mechanical aptitude, following directions in adjusting apparatus, and tracing mazes somewhat analogous to tracing electrical circuits [31]. These tests were given to men in a telephone school

and when properly weighted had a validity of .68. The highest correlation of any one test with the criterion was .52.

SALESMEN

Development of objective procedures for selecting salesmen has proved to be one of the more difficult tasks for the personnel psychologist. This is doubtless due to the lack of good objective tests of personality, a characteristic which seems to be particularly important in this vocation. Tests have shown some validity, but other procedures are also under investigation. Statistical studies have been made of items in the application or personal history blanks and rating scales have been developed for use during the employment interview. Some of these procedures will be described in later chapters. The present discussion will deal with actual tests as applied to selecting salesmen.

One study determined the validity of a large number of individual items [21]. The number of items was as follows: mental alertness, 247; business information, 164; social intelligence, 101; dominance-submission, 29; social attitude, 25; personal inventory, 257; interest, 62. The criterion included rating scales, percentage of the quota sold, percentage of the dealers sold, selling cost per lot. The data were evaluated not in terms of correlation coefficients but rather by noting for certain responses in the test the percentage above or below average in the criterion. In general, alertness (intelligence) proved to be more important for persons in sales promotion than in routine selling; personality items were more valid for salesmen than for sales managers; routine salesmen were more conventional in responses to the social attitude items than were men engaged in promotion.

Another project dealt exclusively with sales managers [24]. The tests were evaluated with a group of successful and unsuccessful managers and six of them had some validity. They included vocabulary, alertness, free association, form completion, giving the names of items which begin with a certain letter, and discovering "pictures" in ink blots.

One other project may be cited by way of transition to the discussion of personality tests [32]. It involved ascendance-submission, introversion-extroversion, two intelligence tests, Strong's Interest Inventory, and a sales objections test. The prin-

cipal result was that a personality test when compared with sales record would have hired 69 per cent of the best men and rejected only 31 per cent. The intelligence test contributed little except in eliminating some of the worst salespeople. This finding suggests the importance of personality in contrast with mental capacity as a factor in selling ability, a point which is being increasingly realized. Attention is called, however, to a point made earlier (p. 112), that in the employment situation considerable caution is necessary in using the type of personality test in which the subject makes statements about himself—what he would do in certain circumstances, what he worries about, or what he prefers. There is always the danger that he will try to discover the best way to answer the item rather than the way which actually characterizes him. In spite of this limitation there have been a number of efforts to standardize tests of this sort for the selection of salesmen. A few of these will be described, although it is hoped that ultimately more objective tests for measuring these same characteristics will be available.

One such study employed the Bernreuter scale (p. 110) with 75 department store salespeople. Managers selected about 20 at each extreme in sales ability [9, 10]. These two groups showed no distinct relationship with any standard Bernreuter score. The correlations were roughly .16 for men and .13 for women. However, an item analysis was then made with a view to picking out the most differential items, and a simple scoring key was derived which did make some differentiation between the two extreme groups. When this key was employed with another sample of salespeople who were rated by the personnel manager the correlation was about .60 for men and .36 for women. The valid categories were the following, the number after each indicating the number of items of that sort included: not moody or subject to worry (7); self-confident (6); self-sufficient (6); social (5); free from self-consciousness (4); aggressive and willing to assume responsibility (3); little tendency to talk about self (3); not resentful of criticism or discipline (2); radical and unconventional (1).

Another test somewhat similar to the Bernreuter and comprising 125 items was administered to 64 salesmen [19]. Their scores were compared with those of 1000 unselected college

students. Standard scoring patterns had been developed previously to indicate a number of aspects of personality. Statistically significant differences between the two groups were noted. The salesmen proved to be less neurotic, to have greater self-esteem and greater independence, and possibly to be a little more extroverted. In another brief check 25 drugstore salesmen were rated by the owner of a chain on a 10-point scale. The lowest fifth in the ratings were decidedly more neurotic and introverted on the personality scale.

The foregoing studies suggest that even personality tests of the sort where a subject does evaluate himself may have some validity in spite of the limitations mentioned previously. Much depends on the situation in which the tests are administered and the explanation that goes with them. In many of the studies reported they were given not as part of an employment program but at some subsequent time and presumably the explanation was that the tests were being standardized, so that the subject did not feel that his own status would be influenced appreciably by what he did on the test. Under these circumstances one might be inclined to answer more frankly than he would if he were actually in an employment office trying to get a job. The results are none too convincing as to the validity of the subjective type of personality test in the practical employment situation.

EXECUTIVES

One investigation of tests for executives made a detailed item analysis [41]. The subjects were about 100 supervisors and foremen in a large manufacturing organization; and special attention was given to the criterion, which consisted of ratings by the subjects' immediate superiors. Each superintendent rated a different group of men, but five persons were rated who were known to all the superintendents and served as "key" men for readjusting the ratings. The estimates were made by ranking and by a graphic rating scale. The two halves of the rating data correlated to the extent of .80. The tests included alertness, mechanical aptitude (paper form board), a personality schedule, and company information. There were 820 items in this group of tests. They were studied item by item, the men being divided into three groups on the basis of the criterion and the percentage

of each group that passed a given item being noted. In this way 85 differential items were selected which showed significant differences between the percentages in the criterion groups. They were a miscellaneous set of items that one would have been unable to select on logical grounds. The company information tests appeared to have a somewhat higher proportion of good items than the other tests. The scores based on the 85 good items correlated with the criterion to the extent of .71. When all the 820 items were combined, the correlation was only .49. This shows what can be done by item analysis in contrast to using gross scores.

A review of a number of existing tests for executive ability may be mentioned [22]. Executives generally score high in general information, reasoning, speed of judgment, detecting symbolic relationships such as analogies and word comparison. They are above average in many qualities; in other words, they are "well-rounded" individuals. The same authors present a suggested test of executive ability along much the foregoing lines, but indicate that certain modifications would be desirable to increase its validity [6]. The test includes general information, personality schedule, reasoning, judgment (for example, "How many horses are there in the United States?"), an analogy test involving geometrical figures, synonyms and antonyms, and some interest items like those in Strong's questionnaire. Further investigations of intelligence as related to executive ability will be presented in Chapter X.

FACTOR ANALYSIS

The technique of factor analysis has been mentioned elsewhere, but it is well to point out its implications for the construction of batteries of aptitude tests. The procedure may be contrasted with that of multiple correlation. In the latter we are concerned with predicting a variable on the basis of a number of other variables, usually trying to predict a criterion on the basis of one or more tests. In factor analysis, however, we are attempting to see if the variables or tests are related by some underlying order that will simplify our comprehension of the whole set. In other words, we are trying to discover the underlying functional unities. As mentioned earlier, Thurstone found

with some fifty-odd pencil and paper tests that apparently nine main factors were basic, such as spatial, visual, numerical, verbal, and so on. It is to be noted that he started with 56 tests and ended with nine factors. This is typical of factor theory. We start out with an array of intercorrelations with as many columns as we have tests, and we finish with one column for each factor; normally these columns will be much fewer in number. When it comes to interpreting the factor loadings we have to use our "hunches" and attempt to identify the factors in the light of these loadings. The procedure is distinctly exploratory.

Thurstone gives an interesting analogy which is too complicated for detailed presentation here [38]. He postulates a set of boxes with twenty measurements for each box. One measure, for instance, might be the square of one edge of a box, another the length of the diagonal of a side, and another the length of the diagonal through the center of the box. We do not know how these measures were obtained; we merely have the gross figures. Nevertheless, some of these measures would correlate with each other. With intercorrelations between the measurements available we have something analogous to the intercorrelations between a group of tests. We now make a factor analysis; the first factor has rather heavy loadings for a number of the twenty measurements on the boxes. By looking at the nature of these measurements we may discover that the height of the box seems to be involved in most of them, so it appears that the first factor is height. Similarly we may find that the depth of the box is represented in another group of measurements for which the second factor is heavily loaded. In this way we may get back to the fact that the three basic factors actually are the height, breadth, and length of the box. Just as one of the initial measures such as the diagonal of a side is a function of the two basic factors of height and breadth, so a particular test may be a function of basic factors such as verbal ability and numerical ability.

A practical result of factor analysis is that we may be able to get at the basic factors involved in a group of tests which show considerable validity and then construct a more limited number of tests aimed directly at those particular factors which will do the job better than the original ones. In other words, we may eliminate some deadwood in the tests. This approach has con-

siderable promise in these problems of aptitude measurement with a battery of tests.

FOLLOW-UP PROCEDURE

After a psychologist has developed a test or series of tests for predicting aptitude in a certain occupation, his task is not completed as far as that occupation is concerned. The methods can be put into use in the employment office for selecting workers. It is desirable, however, to keep a record of the workers' scores in the tests taken at the time they are hired and subsequently to compare these scores with their ability in the job after they have learned it. After a sufficient number of applicants have taken the tests and have been at the job long enough to reach their maximum efficiency, it is well to secure figures as to the latter ability in much the same manner that the criterion was determined originally and then to compare the original test scores with this new criterion. This will serve to vindicate the whole procedure, for while it is probable that tests devised originally to differentiate the good from the poor employees will serve likewise in differentiating the good from the poor applicants, it is well eventually to determine empirically if this is the case. Furthermore, an occasional check on the value of the methods is desirable because there may sometimes be changes in the general type of applicants, the methods of training, or even the methods of work that will render the original tests invalid. This follow-up procedure has a further advantage in that it may be possible from time to time to introduce slight changes in method. It may be desirable to give one or two tests in addition to those originally standardized and evaluate these subsequently with a view to including them ultimately in the regression equation—possibly replacing some of the original ones.

It is well for the personnel psychologist to keep in touch with his original work. It is, of course, often necessary to develop methods, make them as objective and fool-proof as possible, and then turn them over to untrained people in the employment office for routine administration. This is not ideal. The technique of mental examination is more reliable in the hands of a person with psychological training. Unforeseen contingencies may arise. Very frequently extraneous reactions which the applicant makes,

quite apart from his actual test performance, are of vocational significance and only the trained examiner can make the most of this "clinical picture." Some of the larger concerns have psychologists permanently attached to the staff—just as they now have chemists, physicists, or engineers—to maintain constant supervision over the personnel and other work that is psychological in character. An industrial concern is, in a way, a psychological laboratory in which the problems are not solved and the methods devised for all time but in which research may well be continuously in progress.

SUMMARY

In devising a set of tests for the mental components of the occupation the preliminary procedure of analysis is similar to that for the test of the total situation. It is then necessary to select and devise tests for the various components that the analysis reveals. Of course, no test measures an isolated mental factor, but this procedure will probably bring better results than selecting tests at random. The more tests used the greater the chances of finding some that have high correlations with the criterion. The number evaluated generally depends on the length of time for which the subjects are available.

The tests selected must be given to subjects and evaluated to determine which to retain and which to discard. Usually some rough correlation technique is adequate to eliminate the worst tests. The remaining ones are then correlated more carefully with the criterion and with each other in order to assign each test its proper weight in the total score. It is not desirable to weight each test according to its correlation with the criterion because some of the tests may measure substantially the same factor while others may involve more independent factors. This overlapping may be ascertained by correlating the tests with each other.

The technique of partial correlation makes it possible to eliminate the effects of this overlapping. By this technique one computes what the correlation of a test and the criterion would be if based on subjects who all had the same ability in the other tests. This shows the intrinsic relation of a test to the criterion and

affords a more adequate weighting for each test than does its original correlation, which takes no account of the overlapping. A consideration of the partial correlation formulae shows that the best test for predicting the criterion is one which has a high correlation with the criterion and a low correlation with the other tests, for this will tend to make its partial correlation with the criterion high.

A regression equation can then be derived which expresses probable vocational ability in terms of the tests. It indicates the weight or constant number by which to multiply each test score so that the weighted sum will give the best possible prediction of the criterion. The weight for a test is roughly proportional to its partial correlation with the criterion, with the other tests kept constant.

The coefficient of multiple correlation is simply the correlation of the weighted sum of the tests with the criterion. This indicates how valuable the tests are for the purpose in hand and shows how much the weighted sum of the tests is superior to any single test. Another technique makes it possible to add the tests to the battery one at a time and note how the multiple correlation is increased by each additional test. The tests can also be given to the subject in order of their importance and, if he fails to reach the critical score in the first one, he can be dropped from further consideration.

Various examples of tests for the mental components of the job were given. Typical projects involving the following types of employees were included: general factory operatives, clerical workers, drivers of vehicles, public service employees, salesmen, and executives.

When a test project has been developed and put into practical use it is desirable to follow up the results for a time and see whether the new employees hired on the basis of the tests actually conform to the prediction. This serves as a subsequent validation of the whole method and also makes possible minor revisions of the tests. If the psychologist is able to keep in touch with the work it is possible to have a continuous program of occasional addition and revision with a view to gradually increasing the validity of the employment methods.

REFERENCES

1. Andrew, D. M. Analysis of the Minnesota Tests for Clerical Workers. *Journal of Applied Psychology*, 1937, 21, 139-171.
2. Bacqueyrisse, L. La Sélection psychotechnique des machinistes et la diminution des accidents. *Travail Humain*, 1934, 2, 442-447.
3. Bingham, W. V. *Aptitudes and Aptitude Testing*. New York, Harper, 1937, 390 pp.
4. Burt, H. E. Employment Psychology in the Rubber Industry. *Journal of Applied Psychology*, 1920, 4, 1-17.
5. Chleusebairgue, A. The Selection of Drivers in Barcelona. *Occupational Psychology*, London, 1940, 14, 146-161.
6. Cleeton, G. U., and Mason, C. W. *Executive Ability, Its Discovery and Development*. Yellow Springs, Ohio, Antioch Press, 1934, 183 pp.
7. Davidson, C. M. Evaluation of Clerical Tests. *Personnel Journal*, 1937, 16, 57-64, 95-98.
8. DeSilva, H. R., and Channell, R. Drivers' Clinics in the Field. *Journal of Applied Psychology*, 1938, 22, 59-69.
9. Dodge, A. F. Social Dominance and Sales Personality. *Journal of Applied Psychology*, 1938, 22, 132-139.
10. Dodge, A. F. What Are the Personality Traits of the Successful Sales Person? *Journal of Applied Psychology*, 1938, 22, 229-238.
11. Drill, R. Eignungsprüfung für die kommunal Feuerwehr. *Industrielle Psychotechnik*, 1927, 4, 289-301.
12. Garrett, H. E. *Statistics in Psychology and Education*. New York, Longmans, 1937, 493 pp.
13. Gow, C. R., *Elements of Human Engineering*. New York, Macmillan, 1932, 169 pp.
14. Hales, W. M. Clerical Tests in a State Reformatory. *Personnel Journal*, 1938, 16, 316-324.
15. Hall, O. M. An Aid to the Selection of Pressman Apprentices. *Personnel Journal*, 1930, 9, 77-81.
16. Harrell, W. Validity of Mechanical Ability Test for Selecting Cottonmill Machine Fixers. *Journal of Social Psychology*, 1937, 8, 279-282.
17. Hayes, E. G. Selecting Women for Shop Work. *Personnel Journal*, 1932, 11, 69-85.
18. Hooker, R. H. The Correlation of the Weather and the Crops. *Journal of the Royal Statistical Society*, 1907, 70, 1-42.
19. Husband, R. W. Personality Traits of Salesmen. *Journal of Psychology*, 1937, 1, 223-233.

20. Kelley, T. L. *Statistical Method*. New York, Macmillan, 1923, 390 pp.
21. Lovett, R. F., and Richardson, M. W. Selecting Sales Personnel. *Personnel Journal*, 1934, 12, 248-253.
22. Mason, C. W., and Cleeton, G. U. Measuring Executive Ability. *Personnel Journal*, 1935, 13, 277-279.
23. Miles, G. H., and Vincent, D. F. The Institute's Test for Motor Drivers. *Human Factor*, 1934, 8, 245-257.
24. Mitchell, J. H. An Experiment in the Selecting of Sales Managers. *Occupational Psychology*, London, 1938, 12, 308-318.
25. Moore, H. The Institute's Clerical Test in America. *Human Factor*, 1933, 7, 407-409.
26. Münsterberg, H. *Psychology and Industrial Efficiency*. Boston, Houghton Mifflin, 1913, 316 pp.
27. O'Rourke, L. J. Saving Dollars and Energy by Personnel Research; an Investigation in the Interests of the Postal Service. *Journal of Personnel Research*, 1926, 4, 351-364.
28. Otis, J. L. The Prediction of Success in Power Sewing Machine Operation. *Journal of Applied Psychology*, 1938, 22, 350-356.
29. Patterson, D. G., Schneider, G. G., and Williamson, E. G. *Student Guidance Techniques*. New York, McGraw-Hill, 1938, 316 pp.
30. Pond, M. What Is New in Employment Testing? *Personnel Journal*, 1932, 11, 10-16.
31. Rossett, N. E., and Arakelian, P. Test Battery for the Selection of Telephone Dial Switchmen. *Journal of Applied Psychology*, 1939, 23, 358-366.
32. Schultz, R. F. Test-selected Salesmen Are Successful. *Personnel Journal*, 1935, 14, 139-142.
33. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
34. Stedman, M. B. Factors Influencing Success in Bookkeeping. *Journal of Applied Psychology*, 1930, 14, 74-82.
35. Telford, F., and Moss, F. A. Suggested Tests for Patrolmen. *Public Personnel Studies*, 1924, 2, 112-145.
36. Thurstone, L. L. Mental Tests for Prospective Telegraphers. *Journal of Applied Psychology*, 1919, 3, 110-117.
37. Thurstone, L. L. The Civil Service Tests for Patrolmen in Philadelphia. *Public Personnel Studies*, 1924, 2, 1-5.
38. Thurstone, L. L. Current Issues in Factor Analysis. *Psychological Bulletin*, 1940, 37, 189-236.
39. Tiffin, J., and Greenly, R. J. Employee Selection Tests for Elec-

- trical Fixture Assemblers and Radio Assemblers. *Journal of Applied Psychology*, 1939, 23, 240-263.
40. Treat, K. Test for Garment Machine Operators. *Personnel Journal*, 1929, 8, 19-28.
 41. Uhrbrock, R. S., and Richardson, M. W. Item Analysis. *Personnel Journal*, 1933, 12, 141-154.
 42. Viteles, M. S. The Human Factor in Substation Operation. *Personnel Journal*, 1929, 8, 81-113.
 43. Voronov, N. N. Psycho-Technical Examination of Students in a Preparatory Building Trade Course in the Kiev Section of the Central Institute of Labor. *Psikhotekn i psikhofiziol. truda*, 1930, 3, 143-171.

Chapter X

INTELLIGENCE AND VOCATIONAL APTITUDE

The two preceding chapters discussed tests for special mental capacity in so far as they may be used to predict occupational success. The majority of the employment problems with which a psychologist deals are of this type. Occupational misfits are usually lacking in some of these special respects. With modern industrial organization most of the jobs necessitate the acquisition of a relatively small number of habits; and it is a question of whether the applicant has the special capacities such as memory, attention, or quick reaction time, that will facilitate the formation of those habits. There are other cases, however, in which the job apparently does not call for such specialized mental equipment, but rather for an all-round ability, a general mental alertness, or a facility in adapting oneself to a new situation. The characteristic involved here has usually been termed intelligence, and various tests have been devised to measure it. As stated in Chapter IV, it does not matter whether this general ability is called intelligence or something else, and its exact nature is of little consequence. If the results of these general tests enable us to predict occupational success, this is all that is required. The present chapter will be devoted to the use of intelligence tests for predicting vocational aptitude.

OCCUPATIONAL HIERARCHY

Occupational Studies in the Army. One question that arises in connection with such tests in employment psychology is whether a certain minimum of intelligence is required for different occupations. It seems plausible that a person will in the long run rise about as high on the occupational scale as his intelligence warrants, and that, if we determine the average intelligence of people in a certain occupation, this will tell us something about

the general ability required for that occupation. Data bearing on this point were available as a result of the Army Alpha test being given to a large number of men in 1918. In connection with their examination a record was made of their previous occupation. It was a simple matter then to select a group of laborers, or a group of machinists, or a group of professional men, and compute the average intelligence of each occupational group. The results are rather illuminating [18, 819]; a typical portion of them is shown in Table 20.

TABLE 20. INTELLIGENCE OF OCCUPATIONAL GROUPS

	First Quartile	Average	Third Quartile	Per Cent in Class A or B
Engineer officer.....	144	162	176	96
Medical officer.....	117	129	152	77
Civil engineer.....	99	117	143	68
Accountant.....	98	117	136	68
Stenographer or typist...	93	115	138	62
Mechanical draftsman...	84	114	139	59
Mechanical engineer.....	73	110	137	47
Bookkeeper.....	77	101	127	46
Filing clerk.....	74	97	126	40
General clerk.....	74	96	121	40
Railroad clerk.....	69	91	115	37
Telegrapher.....	61	85	110	28
Telephone operator.....	57	70	109	20
Auto assembler.....	51	68	97	18
General mechanic.....	48	68	94	14
Toolroom expert.....	50	67	92	9
General auto repairman..	43	65	91	13
Telephone lineman.....	43	64	88	12
General carpenter.....	40	60	84	9
Baker.....	40	59	87	11
Bricklayer.....	37	58	88	11
Truck driver.....	37	58	83	11
Barber.....	34	55	78	7
Boiler-maker.....	31	51	74	9
Teamster.....	30	50	72	6
Miner.....	40	49	71	5
Farmer.....	30	48	73	7
Laborer.....	28	47	68	4

The scores are in terms of the actual number of points made out of a possible 212. The first column gives the first quartile score or 25 percentile, i.e., the score which one-fourth of the group fails to surpass. The next column gives the average. The third one gives the third quartile or 75 percentile, i.e., the score which three-fourths of the group fails to surpass. The distance between the first and third quartiles obviously includes the middle half of the group, and it is often used as a rough measure of variability. If the first and third quartiles are close together, this indicates that the individuals are "bunched" or have a small variability. The Army test also utilized letter grades, C being average intelligence, B high average, and A superior intelligence. The last column of the table gives the percentage of each group in Class A or B.

The table includes only a few of the many occupations that were studied in this fashion, but it is sufficient to afford an idea of the general trend. There is definite evidence of an occupational hierarchy. At the bottom of the intelligence scale are the unskilled laborers; higher up are those in more skilled mechanical occupations; above those are the clerical and business workers; and at the top those in the professions. It seems reasonable that the intelligence requirements of the professions should be more exacting than those of the unskilled laboring jobs and that the figures should give some indication of the intelligence necessary for the occupations in question.

Extension of Original Studies. A considerable extension that attempts to cover the whole gamut of occupations has been made of this earlier study of the occupational hierarchy. Only one of the more extensive of these later investigations will be cited. It was made at the University of Minnesota and represents the pooled judgment of twenty industrial and vocational psychologists [5, 366]. Six categories of occupations are listed in the order of the minimum necessary intelligence.

1. High professional and executive occupations. These require superior intelligence and comprise ability for creative and directive work such as president of a college or a large manufacturing concern. Graduation from a first-class college is also requisite.

2. Lower professional and large business occupations. Superior

intelligence and training equivalent to two or three years of college are required. Persons in this category need be less creative than those in Group 1, but they exercise leadership such as executive of a moderately large business, or high school teacher.

3. Technical clerical supervisory occupations. These require high average intelligence and at least high school graduation. They include minor executives and people doing highly technical work dealing with abstract details such as railroad workers, some retail dealers, shop foremen, stenographers.

4. Trades and low-grade clerical jobs. They require average intelligence and the equivalent of some training beyond the eighth grade. They do mechanical work which demands specialized skill and knowledge, and tasks of a complicated but concrete nature, especially those requiring technical training like auto mechanics, filing, typing.

5. Semi-skilled occupations. These require low average or slightly below average intelligence and training equivalent to the eighth grade. The work demands a minimum of technical knowledge or skill but certain special abilities like dexterity.

6. Unskilled occupations. Only inferior intelligence is required; no formal training is necessary. Routine manual work is done under supervision, such as laborer.

It is interesting to note that a similar hierarchy is obtained when we consider merely the intelligence of children compared with the occupations of their parents. Such a study of 1200 school children yielded the following average intelligence quotients for the pupils whose parents were in the occupational classes indicated: professional, 113; executive and business, 107; skilled labor and clerical, 97; semi-skilled labor, 84; unskilled labor, 76. While there was much overlapping, this hierarchy on the part of the children's intelligence as related to their parents' occupation is interesting. It assumes, of course, that intelligence is hereditary [12].

Specific Occupational Groups. Several studies have been reported in which intelligence tests were given to miscellaneous specific groups of workers and the averages noted in the fashion discussed above [21]. The results of such a study are given in Table 21. The figures indicate the average scores in a test in

which the maximum possible score is 113 points. The results are quite similar to those obtained with the Army tests. The

TABLE 21. AVERAGE INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS¹

College presidents (small colleges).....	58
Engineering students.....	57
Students in medical college.....	56
Students in arts college.....	54
Sales executives.....	54
Supervisors in manufacturing plant.....	52
Executives of progressive firms.....	51
Rotary Club members.....	46
Nurses.....	42
Foremen.....	41
Office employees.....	40
Machine operators.....	33
Sales force in department store (men).....	33
Office boys.....	31
Sales force assisting in holiday rush (men).....	29
Students in business college.....	28
Sales force in department store (women).....	27
Sales force assisting in holiday rush (women).....	25

retail salespeople have the lowest intelligence of those studied. Machine operators, office employees, and foremen are somewhat superior. Rotary Club members, who are presumably successful business men, are higher still. Executives and college students make still better scores, and college presidents are at the top. This hierarchy does not extend as far as the unskilled laborers.

Results of a similar investigation appear in Table 22 [15, 275]. A few of the groups included are rather selected and hence score higher than would similar groups taken at random. This is particularly true of the major executives, the first-year graduates in business school, and the real estate salesmen. The retail clerks likewise are confined to a group selling a particular class of commodities and a more random selection would probably make lower scores. The hierarchy, however, is rather obvious, ranging from major executives, engineers, and students down through school teachers, general executives, special groups of salesmen, office workers, routine salesmen, and policemen, to the retail clerks at the bottom of the intelligence scale.

¹ After Scott and Clothier.

TABLE 22. INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS²

	First Quartile	Average	Third Quartile
Major executives	90	127	156
First-year graduates (business)	109	125	140
Sales engineers	110	120	150
College seniors	100	118	137
School superintendents and special subject teachers	100	109	119
Executives (general)	82	102	116
Real estate salesmen	80	102	115
Office specialty salesmen	60	95	112
Students in school for insurance salesmen	74	93	105
Experienced insurance salesmen	60	86	110
Office clerks	55	84	105
Semi-specialty salesmen	56	78	106
Routine salesmen	41	71	94
House-to-house salesmen	30	65	95
Trade high school (night)	37	62	87
Policemen	a	42	a
Retail sales clerks (notions, bar- gain counters, etc.)	20	33	50

a Figures not available.

Academic Level. A notion of the functioning of this hierarchy at the upper end of the intelligence range may be obtained from a consideration of the results of the Army Alpha test given to the entire student body of a large university and a majority of the members of the faculty who taught those students. The average for the students is 136 points actual score in the test, and for the faculty 154. The average for the entire Army was about 60 points. The middle half of the student scores fall between 116 and 155, while the middle half of the faculty scores fall between 139 and 174. The figures are expressed in terms of the conventional letter grades in Table 23. A means very superior intelligence, and C means average. The table gives the percentage of faculty and of students in each of these classes

² From H. G. Kenagy and C. E. Yoakum, *The Selection and Training of Salesmen*, by permission of the McGraw-Hill Book Company, Inc., New York.

and also, for comparison, the percentage of soldiers falling in these classes. This table shows the superiority of the faculty to the students, and in turn the superiority of both to the unselected men in the Army, who probably represented the whole range of occupations.

The results should be somewhat qualified by the fact that the examination was voluntary with the faculty but compulsory with the students. It has been found in other connections that

TABLE 23. PERCENTAGE OF FACULTY AND STUDENTS IN DIFFERENT INTELLIGENCE CLASSES

Class	Faculty	Students	Army
A	77	51	5
B	16	33	10
C+	4	13	18
C	2	2	29
C-	0.4	0.2	21

persons willing to take an intelligence test, or at least willing to write their names on their test blanks, grade somewhat higher than those reluctant to do so. This would tend to lower the faculty results to some extent if the entire group had been involved. However, with differences of the magnitude shown in the table, there is clear indication that those in the profession of college teaching stand higher in the intelligence hierarchy than their students, many of whom will ultimately settle in much less intellectual types of occupation.

Another study at the academic level dealt with over 1100 alumni of a small college [11]. They had all taken an intelligence test while they were in school. The results are based on intelligence percentiles which were established each year with the students who took the test. The results are summarized in Table 24. The subsequent occupations of the alumni are arranged approximately in the order of the intelligence of those included in the particular group. The columns give the average and quartile scores of the various occupational groups. A considerable overlapping of the groups is obvious from a consideration of the quartile points, but nevertheless there is indication

of a hierarchy similar to that mentioned previously, ranging from college teaching down to physical education.

TABLE 24. INTELLIGENCE AND OCCUPATIONAL GROUPS³

	First Quartile	Average	Third Quartile
College teaching—women.....	49	77	89
College teaching—men.....	43	65	83
Secretarial work.....	41	59	73
Journalism.....	41	57	85
High school teaching—men.....	24	57	75
Medicine.....	33	54	78
High school teaching—women.....	26	53	71
Religious work.....	18	51	78
Law.....	20	51	71
Music.....	24	50	73
Business—men.....	25	50	72
Other educational work.....	24	49	72
Engineering and scientific work.....	21	49	75
Business—women.....	32	49	58
Library work.....	21	46	82
Salesmanship.....	16	44	71
Social work.....	22	41	73
YMCA and YWCA.....	26	39	71
Physical education—women.....	24	35	50
Art.....	20	27	70
Physical education—men.....	9	20	31

Different Departments in an Organization. A set of tests, the combination of which was tantamount to an intelligence test, was given in a rubber tire plant. The average scores of various occupational groups within this one concern appear in Table 25. The highest scores are attained on the average by a group of employees in the laboratory and drafting departments. These individuals are, of course, technically trained. Slightly inferior to them, but perhaps not significantly so, are the members of the factory council, a group of six executives who at that time determined the policies of the organization. Below these come a group of general clerical workers, who compare rather favorably

³ After Hartson.

TABLE 25. AVERAGE INTELLIGENCE SCORES OF OCCUPATIONAL GROUPS IN ONE CONCERN

Laboratory and drafting.....	147
Factory council.....	144
General clerical workers.....	138
Shipping department.....	112
Factory committee.....	108
Foremen.....	88
Inspectors.....	86
Finishers and builders.....	87
Handing out stock.....	76
Truckers and mixers.....	47

with the executives and are distinctly superior to the other groups involved. The next in order are the employees in the shipping department, followed closely by members of the general factory committee. This committee was comprised of a few foremen and various minor executives who met regularly to determine less important questions of policy. Between this group and the foremen and inspectors there is a considerable gap. The men who finish and build tires compare favorably with the foremen and inspectors. This probably reflects the well-known fact that foremen are chosen in some concerns not by virtue of any superior capacity, but simply because they are experienced workmen. The employees who hand out stock are somewhat inferior to the finishers and builders and foremen. Far down at the bottom of the scale are the employees engaged in unskilled labor, such as hauling trucks or mixing and washing crude rubber.

Different Types of Salesmen. The foregoing discussion has dealt with the occupational hierarchy for the whole range of occupations from unskilled labor to the professions. The question arises whether there is any such hierarchy within a given occupation. Some data for salesmen are available on this point. A number of the occupational groups listed in Table 22 (*supra*) may be classed as salesmen and there is some evidence of a hierarchy. The sales engineers make the highest scores in intelligence. The real estate salesmen are appreciably lower. A little lower still are the office specialty salesmen and the students in a school for insurance salesmanship. The experienced sales-

men are inferior to the students, but the two groups manifestly overlap considerably. Next in order come the semi-specialty salesmen, then the routine salesmen, with the house-to-house salesmen lower still. At the bottom of the scale, far inferior to any of the others in intelligence, are the retail sales clerks.

Quite similar results were obtained in another study of four different groups of salesmen [17]. The results are presented in Table 26. The same tendency is manifest. The men who sell a

TABLE 26. INTELLIGENCE SCORES OF GROUPS OF SALESMEN⁴

	First Quartile	Average	Third Quartile
Salesmen for technical product . . .	124	139	155
Insurance salesmen	82	112	138
Wholesale salesmen	59	89	121
Counter salespeople	36	51	70

highly specialized technical product stand at the top, and the counter salespeople are at the bottom. There is considerable overlapping, especially of the wholesale and insurance groups, but there is sufficient difference to be of interest.

It would appear that even within a single vocation, like selling, there is an intelligence hierarchy. All salespeople have considerable in common in that they are inducing prospects to purchase something. But it seems that even with this common element certain types of selling actually require a higher order of intelligence than do others. These results should not be interpreted to mean that intelligence tests alone will give a good prediction of selling ability. Neither do they imply anything about the diagnostic value of intelligence tests within a particular group of salespeople, such as retail clerks. They do indicate, however, that over and above the other mental qualifications requisite for salesmanship certain aspects of this vocation are more exacting in their intelligence requirements than are others.

The theory underlying the various results just presented is

⁴ After Miner.

that a person will in the long run tend to rise about as high in the occupational scale as his intelligence warrants. If he attempts a job too high in the scale, he will find it too exacting and leave either voluntarily or involuntarily. If, on the other hand, he starts with one that is too low in the scale, he will not find it sufficiently interesting because it does not afford an adequate outlet for his intellectual ability, and he will leave it for something higher. The result is that he ultimately lands at about the maximum level at which he can do effective work. Other factors, of course, sometimes alter the results. A lazy individual may not want a more exacting type of work, or a person unattractive in appearance or with some personality defect may be refused a job for which he is capable. The above assumption deals only with the average case. If it is valid, we may conclude that the average intelligence of an occupational group indicates approximately the degree of intelligence that is necessary for effective work in that occupation.

These principles may be used to some extent in the practical problem of employment. If it is known, for instance, that persons below a certain intelligence score are seldom found in clerical or executive positions, it will probably be well, in lieu of further special examinations, to select for such positions people whose intelligence is at least equal to the critical amount. The lines cannot be drawn too closely, but extreme values surely are significant. Persons of very low intelligence, such as that possessed by the average unskilled laborer, would doubtless be distinctly misplaced when put in an executive or clerical position, and it would presumably be better policy to give them unskilled or semi-skilled laboring work at the outset. By this procedure no one can hope to predict an individual's success in a given line of work in terms of probability, as is possible when a correlation coefficient is available. The most that can be done is to locate the individual somewhere near his appropriate occupational level. This information, however, is often valuable, especially when dealing with extreme cases of discrepancy between the intelligence possessed by the applicant and that required for a given occupation. Incidentally, this occupational hierarchy is used in vocational guidance perhaps more than in vocational selection. The counselor on the basis of intelligence tests can locate the

individual somewhere near the level where he might hope to be successful, and advise him accordingly.

VOCATIONAL POSSIBILITIES FOR THE FEEBLE-MINDED

A question of some industrial and certainly of some social significance is the possibility of utilizing persons who are mentally low-grade or even defective. Cases are reported in which an industry, particularly one in a small town, is able to use some of the low-grade individuals to the benefit of all concerned [3]. For instance, among some hundred other workers in a knitting mill there were 24 morons who had filled in when some of the others migrated to the city. Their output was about 55 per cent of normal production and they were paid accordingly. An additional advantage was that they did not lose much time through diversion by outside interests. Such a program implies that the cases are carefully selected and that their outside life is reasonably well supervised. In many instances it is inadvisable to have morons running loose in the community because of their tendency to get into trouble, and even delinquency.

Some investigations have determined the minimum intelligence or mental age that could be utilized in industrial work [8]. In one study the minimum was placed at a mental age of $6\frac{1}{2}$ years. Such persons could clean mirrors in pocketbooks and pack powderpuffs. People with a mental age of 7 to 8 could assemble simple electric parts, put buttons on cards, and do operations involving pasting. Between the mental ages of 8 and 9 people could address envelopes, do bottling, seam dresses, and work on feathers and artificial flowers. Another study gives similar age levels for different types of work and also includes data from a laundry in which persons with a mental age of 9 did satisfactory work [2]. In another report dealing with the garment trade, 84 subnormal girls with an average I.Q. of .66 were followed up for a period of $3\frac{1}{2}$ years. A significant factor in their case was that they were willing to stick to routine monotonous jobs and had no aspirations to advance beyond that level. As suggested earlier, it is possible for a person to be too intelligent for a job so that he becomes dissatisfied and inclined to leave. Other factors contributing to the successful adjustment of individual cases were stable homes, careful job placement, encouragement, and

patient treatment during the initial work period [1]. Occupational adjustment for the borderline or mental defective is of considerable social importance. Emphasis should be given again, however, to the necessity for adequate social supervision of such individuals when they are outside of the industry.

INTELLIGENCE SURVEYS

It is sometimes profitable to conduct a survey with intelligence tests throughout an organization or a group of similar organizations. This often reveals conditions that were unsuspected and that will throw rather interesting light on employment problems. Quite apart from devising methods of predicting occupational success, it is frequently of interest to determine what progress has been made to date with the usual employment methods. If the survey is conducted on a rather large scale, sampling a considerable range of jobs within the plant, it is probable that the usual hierarchy will be found, as was the case with the concern surveyed in Table 25. Other factors, however, are sometimes brought out in such surveys; a few typical results are given below [16]. One cannot tell in advance just what to expect, but often something will turn up that will throw rather interesting light on employment problems.

Male vs. Female Employees. A company that had a large number of office employees of both sexes compared the intelligence of the two groups. In the particular test used, the male office employees averaged 51 points and the female employees averaged 38 points. This casts no reflection on the intelligence of women in general. It shows merely that the company had selected for its office a somewhat higher grade of men than of women. It is possible, too, that some men of high intellectual capacity take a clerical position as a steppingstone to executive work. At any rate, the results indicate the desirability of judging male office employees by standards derived from testing men, and vice versa.

Similar Employees in Different Companies. A survey was made of the women office employees in several different companies. In one company their average intelligence was 31 points, in another 38 points, in another 42, and in a fourth 46 points. Obviously the companies had different standards and some were

more exacting than others. A similar situation was found with reference to office boys. In one company their average intelligence was 26, in another 32, and in a third 36 points. Evidently the last company was employing a higher type of personnel for this work. The boys in this concern would manifestly be a better source from which to recruit future executive material.

Applicants vs. Employees. In two concerns where the intelligence of the office employees had been determined, similar tests were given to all the applicants for office work. In the first concern the applicants averaged 36 points in intelligence and the employees 29 points, while in the second the applicants averaged 38 and the employees 47. Evidently both were attracting about the same grade of applicants. The second one, however, was employing a much higher type of personnel. To analyze this difference it would be necessary to know more about the employment methods of the companies and their wage policies. There were preliminary indications that the first company selected high-grade individuals but failed to keep them because they left for more lucrative positions elsewhere. Data of this sort raise the problem of further analysis of the policies and methods of the concerns. An interesting point for investigation would be a comparison of the intelligence of the firm's personnel with that of the general population.

CORRELATION OF INTELLIGENCE WITH THE CRITERION

The foregoing methods are not the only ones by which the problems of intelligence and employment may be approached. The technique described in earlier chapters for evaluating special capacity tests is likewise applicable with reference to intelligence tests. A group of persons engaged in a certain occupation may be given intelligence tests and their test scores correlated or otherwise compared with the criterion. The statistical techniques are exactly the same as those described previously. In some of the cases to be presented, the criterion consisted of production figures or careful estimates made by the employees' superiors, and correlation coefficients were computed. In other cases a less refined comparison was made of different groups of workers.

Clerical Workers. With a group of women office workers the correlation of intelligence and the supervisor's estimate as to the

worker's ability was .76 [16]. With a similar group in another company the correlation was .82. These are comparatively high correlations. In one other group the correlation proved to be only .40. Subsequent analysis revealed, however, that the supervisor had rated the women on the basis of length of service rather than on actual proficiency. It would seem that intelligence is one of the main requisites for this kind of work, or at least that those who have high intelligence possess the other necessary qualifications.

In the preceding chapter we discussed a number of projects for selecting clerical workers. In many of these the tests employed were similar in content to the materials embodied in typical intelligence tests. It is sometimes difficult to draw the line between "clerical" tests and "intelligence" tests. In the latter, it is often a matter of collecting a considerable variety of types of items and the total score is presumed to measure intelligence. It appears that many of these same items are valuable in discovering potential clerical ability. If a distinction between the two kinds of items were attempted, it would probably be to the effect that intelligence tests embody more abstract items whereas clerical tests deal with more specific things like numbers and proper names.

A study was made of clerical workers in two large organizations using a test somewhat similar to Army Alpha [20]. In a life insurance company data were available for about 900 clerical workers, of whom 15 per cent were stenographic. The jobs were graded in five classes from routine to supervisory. Actually the top class (E) presented in the table includes a few persons in a still higher class. The results appear in Table 27. The first column gives the classification of the job, E being the high grade and A the low grade. The other columns give, for the range of intelligence test score indicated at the head of each column, the percentage distribution in the grades of jobs. For example, with individuals scoring 80 points or less in the test, only 6 per cent are in the E class of clerical work, 20 per cent in the C and D classes, and 74 per cent in the A and B classes. It is obvious that with increased intelligence there is an increasing proportion in the higher type of clerical jobs. This is tantamount to a high correlation between the two variables.

TABLE 27. INTELLIGENCE AND GRADES OF CLERICAL WORK⁵

Job	Test Score		
	0-80	81-100	100+
E and better.....	6	12	29
C and D.....	20	29	33
A and B.....	74	59	38

In a metal trade establishment the clerical force was rated from Class 1, which included office department head, district office manager, sales supervisor, and assistant, down to class 7, checkers. A tabulation was made of these seven classes against grades in an intelligence test similar to that just mentioned, and the correlation was .58 for a sample of 286 men and .39 for a slightly smaller sample of women. Both of these studies assume that the clerical worker finds his appropriate level. Those with good ability who begin in a rather low-grade position will in the long run tend to be promoted and those who start at too high a level are apt to be demoted.

Office Boys. Among a group of messenger boys, those discharged averaged 22 points in an intelligence test, while those promoted averaged 39 points. In another case a group was tested and the results were filed for twenty-one months. At that time the average score of those who were still in the company was 42, and of those who were not in the company 35. Further analysis of the latter group revealed that those who left to accept better positions averaged 45 points and those who were discharged averaged 28 points. Of those who remained, the ten boys who stood highest in the test were receiving an average salary of \$16, and the ten who were lowest were receiving an average of \$13.40 [21, 266]. The executives under whom these boys worked estimated their future value to the company by classing them into four groups as follows:

- A. Probable high-grade executive ability.
- B. Probable minor executive ability.

⁵ After Pond and Bills.

- C. Without executive ability, but good clerical timber.
- D. Probably best adapted for highly mechanical job.

The average intelligence and the average salary of those in each of these four classes is given in Table 28. It is to be noted that

TABLE 28. INTELLIGENCE OF OFFICE BOYS⁶

Executive's Estimate	Average Salary	Average Test Score
A.....	\$16.78	66
B.....	\$14.48	55
C.....	\$13.74	51
D.....	\$14.26	39

the executive's ratings and the test scores agree perfectly. There is likewise a fair agreement between salary and the other two factors.

Clothing Operators. The production of operators in clothcraft shops was correlated with intelligence. The coefficient was .51. In using the test subsequently some persons with low scores were hired but were assigned to less exacting work. The conclusion was drawn that "in clothcraft shops the use of mental tests, although only a partial measurement, is the quickest, most accurate, and most economical method of prophesying future skill at machines and of placing operators at types of work suited to their capacity" [21, 266].

Executives. An intelligence test was given to minor executives and five years later the results were compared with their firm rank. The correlation was .69. A small group of executives at the head of a concern were ranked by the vice-president as to their executive ability. The correlation with their rank in an intelligence test was .89.

When business success in general is considered, rather than executive ability within a single organization, a somewhat different result is obtained. A group of business men at a conference took an intelligence test [6]. They subsequently received a questionnaire dealing with their business career, and on the basis of

⁶ After Scott and Clothier.

these questionnaires five judges rated them as to "success." The judges agreed fairly well among themselves as shown by an average correlation of .60 between the different judges. The combined "success" rating correlated with intelligence to the extent of $-.10$. The following conclusion was drawn: "The evidence in hand suggests that superiority in intelligence above a certain minimum contributes relatively less to business success than does superiority in several non-intellectual traits of personality."

Salesmen. While the foregoing results have indicated in most cases some correspondence between intelligence and occupational efficiency, it is unsafe to generalize and conclude that this is true of all occupations. Many instances are found in which the results are not so clear cut. Salesmanship is one of them. The results are somewhat equivocal, but in general the relation of intelligence to selling ability is slight. With two groups of retail sales clerks the correlations between managers' ratings and intelligence were $-.11$ and $-.26$ [15, 260]. This indicates a small inverse relation between intelligence and the criterion. On the other hand, a group of shoe salesmen were classed by executives as good and mediocre [24]. The former ranged from 33 to 59 in test score, and the latter from 19 to 44. Similarly the saleswomen in the same establishment were rated as above average, average, and below average. The average scores of these three groups were respectively 95, 71, and 41, although there was more overlapping in the groups than in the case of the men.

For house-to-house salesmen a zero correlation was found between production and intelligence. It seemed that a man with low intelligence stood as good a chance of success in this line as did a man with high intelligence [15, 261]. With two groups of routine salesmen the correlations were respectively $-.06$ and $.00$. There was, however, a little indication that those of lower intelligence were better than those of high intelligence. For the men who were above average in production the average score was 64, for those who were average in production it was 65, and for those below average in production it was 78. Similar results were found with heating equipment salesmen. The correlation was insignificant, but the average scores for above-average, average, and below-average salesmen were respectively 74, 72, and

94. This may have been due to the fact that a considerable number of high-grade men had been employed recently and had not had sufficient time to demonstrate their ability.

Of a large group of life insurance men the sales managers averaged 93 points in a test and the whole group of salesmen 83 points. Promotion to managership in this field usually depends on success in selling, so there was some indication of the value of intelligence. In a smaller group the correlation of intelligence with two-year production was .24 and in another group the correlation with four-year production was .34. In a single company the correlation of intelligence and production for a small group was .60.

For the office specialty salesmen in two companies intelligence showed a very slight correlation, but there were some indications of relationship when the managers were considered in comparison with the salesmen. The average intelligence scores of managers, active salesmen, and inactive salesmen in the first company were 76, 73, and 69 respectively, and in the second company 74, 69, and 73. In so far as promotion to the position of manager indicates success, there is a slight indication of a positive relation between intelligence and success in selling this specialty.

These results do not conflict with those presented earlier regarding the intelligence hierarchy. It was shown there that certain types of selling are somewhat more exacting from the standpoint of intelligence than are others. But when salesmen of a given kind are considered, the results are not very clear cut. There is some indication that in the lower grades of selling, such as retail clerking, there is a slight negative relation between intelligence and proficiency, while at the upper end, such as insurance or specialty selling, there is a slight positive relation [14]. The smallness of these relations may be in part due to the fact that salesmanship is in a period of transition from selling through individual efforts to selling through advertising, so that the salesman's work is at present less definite and measurable. Production figures for selling, moreover, are influenced by extraneous factors, such as territory, to a greater extent than are similar figures for workers in a factory. At any rate, it is more difficult

to predict selling ability on the basis of intelligence than it is to predict some of the other occupational abilities above mentioned.

Silk Mill Operatives. A large number of employees in a silk mill were given various intelligence tests, mostly of the performance type. The correlation between tests and production was practically zero. "The best weaver in the mill took 10 minutes to assemble a puzzle that an intelligent person does in 25 seconds" [19]. It seemed that in the work where the machinery was automatic and little skill was needed, high intelligence was not required and might even be detrimental. It is quite possible that a person of high intelligence will revolt at such monotonous work, and that one requires stolidity, patience, inertia of attention, regularity of habits, and other temperamental traits rather than intellectual traits.

Operations in an Industrial School. The boys in various occupational groups at an industrial school were rated in proficiency relative to the others in that same trade [10]. They were given Binet tests and mental age was correlated with trade rating. In most instances the correlation of intelligence with the criterion was small. However, there were a few appreciably positive correlation coefficients, and also a few negative coefficients. Some of these are given in Table 29. Office work shows a very high correlation. This suggests the similar fairly large coefficients mentioned above for clerical workers. The poultry department likewise shows a fairly high coefficient; hospital and printing work follow. On the other hand, a few negative coefficients are manifest. The largest of these is for plumbing and the next in order

TABLE 29. CORRELATIONS OF INTELLIGENCE AND
TRADE ABILITY IN AN INDUSTRIAL SCHOOL⁷

Office.....	.98
Poultry.....	.60
Hospital.....	.41
Printing.....	.33
Gardening.....	-.23
Laundry.....	-.30
Bookbinding.....	-.31
Shoe shop.....	-.31
Plumbing.....	-.38

⁷ After Cowdery.

for shoe shop, bookbinding, and laundry work—all on a par. While these coefficients are not large, their existence is suggestive. It is possible that in some of these types of work greater proficiency goes with lower intelligence—provided proper supervision is given. It is probable that the boys in an industrial school are supervised more carefully than the average adult in industry. Consequently the correlations might not be so large in the usual practical situation.

The obvious implication of these studies of vocational proficiency as compared with intelligence is that intelligence tests are valuable in selecting employees for some kinds of work, but that for other types of work they are worthless. It is an unwarranted assumption that for a particular job the most intelligent person available is to be preferred. Just as in dealing with tests for special capacity it is necessary to test the tests, so, in dealing with intelligence as predictive of ability for a given occupation, it is necessary first to correlate efficiency in the test with efficiency in the job. As far as intelligence tests have been employed in industry, they have proved most useful (aside from locating workers at approximately their appropriate level in the hierarchy of occupations) in selecting clerical workers, office boys, and executives.

CRITICAL SCORES IN INTELLIGENCE

Method. If it is established that intelligence is related to proficiency in a certain job and the tests are to be used for employment purposes, the problem arises of establishing a critical score as a basis for hiring or rejecting applicants. The procedure here is identical with that used in the case of tests of special capacity. The most probable ability in the job can be computed from a regression equation or by the use of distributions like those in Table 16, and the decision made as to how big a chance it is desired to take. Or the critical score can be set by inspecting the data—comparing extreme cases—or by determining in a scatter plot where a line can be drawn with the least overlapping of two classes of vocational ability. A few examples of critical scores determined by one or another of the usual methods will be cited by way of illustration.

Examples. In a large tire manufacturing concern in which in-

telligence tests were rather extensively used, critical scores were established for a considerable number of jobs [21, 242]. Some of them are given in Table 30.

TABLE 30. CRITICAL INTELLIGENCE SCORES IN A TIRE CONCERN⁸

Women:

Stenographers.....	35
Typists.....	33
Comptometer operators.....	27
Clerks.....	23

Men:

Factory school instructors.....	50
Chemical engineers.....	45
Other engineers.....	40
Draftsmen.....	35
Clerks.....	30
Dispatch clerks.....	30
Inspectors and foremen.....	23
Messenger and mail boys.....	15

In the study of office boys mentioned above, it was decided to set a critical score of 32 points. On this basis only 43 per cent of the boys below this score remain with the company, while 62 per cent of those above this score remain. The group scoring below 32 points includes only 1 of the 29 boys who were promoted but all 16 of those discharged.

In the study of shoe salesmen already mentioned, a critical score of 33 points would rule out none of the good group and would eliminate 57 per cent of the mediocre group.

In office specialty selling a score of 50 seemed to be critical. All the managers scored above this. Consequently, in employing prospective managerial material persons above this score were selected. In one company, of the 19 men below this critical score, 7 left the employ, 8 produced very little, and 2 of the remainder were below average in production [15, 265].

In connection with the work of a vocational adjustment bureau [9], critical scores for a number of types of work were determined on the basis of mental age. For instance, in millinery work girls whose mental age was 9 and 10 years seemed adapted to such work as sewing linings in hats or steaming material. A

⁸ After Scott and Clothier.

mental age of 11 was necessary for an improver's job, i.e., an operation in which the foundation of the hat is covered and a wire edge attached. A mental age of 12 seemed necessary for machine work on straw or other material.

OPTIMUM VS. MAXIMUM INTELLIGENCE

It might be supposed that in a given vocation in which there was some relation between intelligence and success, it would be advisable to hire persons with the maximum intelligence. A critical score might be established for the minimum intelligence that would enable one to do satisfactory work, but it might be supposed that above this figure the more intelligence the applicant possessed, the better. Recent work, however, has shown that in some cases there should be an upper critical score as well as a lower. In other words, what is needed is prospective workers not with *maximum* intelligence but rather with *optimum* intelligence. These facts come out clearly in studies of turnover or permanency in relation to intelligence and reveal that a person may be too intelligent for his job so that it fails to interest him and he quits.

Stability and Intelligence: Office Workers. In a survey of an office force, stability was plotted against intelligence [16]. These results are shown in the two upper curves of Fig. 7. Along the base line are the test scores. The vertical distances represent the percentage of those with a given score who leave the job within six months from the time they were hired. The results are most striking in the case of the women clerks. Those with scores between 30 and 50 are more stable than the others. A large percentage of those with low intelligence leave, presumably because they do not have sufficient ability to be effective in this line of work. But a large percentage of those with high intelligence likewise leave. It is probable that the job is not sufficiently exacting to hold their interest. High intellectual capacity apparently demands expression or exercise, and these women are discontented.

Similar results were found in another company. Between 40 and 50 per cent of the women clerks with high or low intelligence left the office within six months, whereas about only half

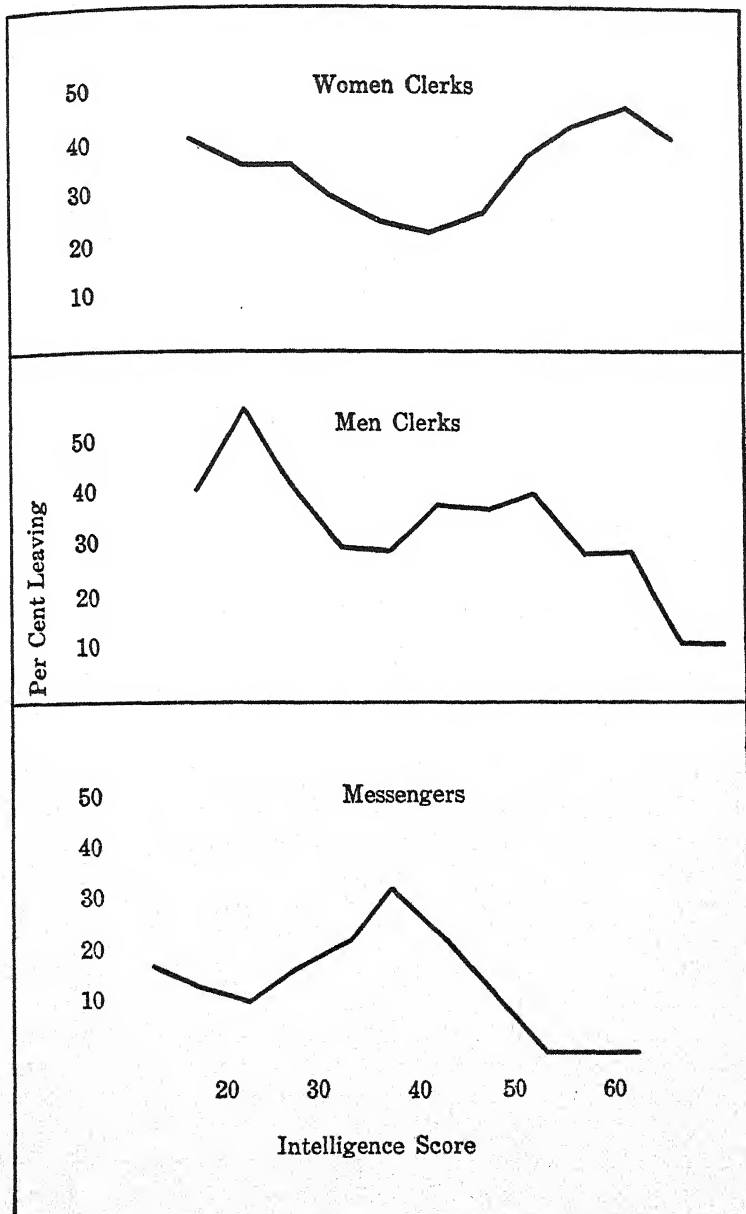


FIG. 7. INTELLIGENCE AND OCCUPATIONAL STABILITY

as many with medium intelligence left in that period. In another case there was a correlation of $-.45$ between intelligence and length of service. This means that the more intelligent worker left earlier than the less intelligent.

In a large clerical force turnover was computed for a period of 30 months [4]. The work was graded in five degrees of difficulty denoted by A, B, C, D, E—A being the lowest grade of clerical work. Two arbitrary points in the intelligence scale were selected—80 points and over 100. The results are shown in Table 31. One notes immediately that for low-grade jobs (A and B) the most intelligent workers have the highest turnover, while for

TABLE 31. TURNOVER FOR CLERICAL WORKERS OF HIGH AND LOW INTELLIGENCE⁹

Grade of Work	Per Cent Turnover for Intelligence Less Than 80 Points	Per Cent Turnover for Intelligence Over 110 Points
A.....	37	100
B.....	62	100
C.....	50	72
D.....	58	53
E.....	66	41

the lowest-grade job (A) the least intelligent are the most stable. In still another company 40 per cent of the clerks who scored less than 30 points in the test left within six months. This percentage decreased up to about 50 points in the test, then increased again for those who made higher scores [22].

Messenger Boys. Results at variance with the foregoing were obtained with a group of messenger boys [21, 253]. They are shown in the lowest curve of Fig. 7. The fewest resignations occurred among the boys with high and low scores. For the low group this was perhaps due to the fact that the applicants were sufficiently alert to hold the job but incapable of improving themselves by going elsewhere. The data do not include boys who were discharged. The results for the high group were explained in this particular case by the fact that the work was not

⁹ After Bills.

distasteful to the brighter boys, because it afforded them an opportunity to learn a good deal about the business and might serve as a steppingstone to a higher position in the office. Many prominent executives have, of course, come up from the ranks of the office boys.

In a parcel delivery service, on the other hand, the turnover was 60 per cent for those scoring below 25 points in an intelligence test, 14 per cent for those scoring between 25 and 47 points, and 75 per cent for those scoring over 40 points [13].

Cashiers. A group of cashiers and inspector wrappers were tested and the results were compared with stability [26]. The facts are shown in Table 32. It is obvious that the greatest stability is found in the middle range of intelligence.

TABLE 32. INTELLIGENCE AND LENGTH OF SERVICE¹⁰

Test Score	Average Length of Service in Days
10 to 19	3
20 to 29	91
30 to 39	156
40 to 49	142
50 to 59	107
60 to 69	100
70 to 79	96
80 to 89	87
90 and over	35

Policemen. The Army Alpha Test was given to a group of policemen in a large city [25]. The average scores for different groups are presented in Table 33. The results shown in the first three lines of the table are not what one would ordinarily expect. One would suppose that the officers would have higher intelligence than the men under them, but this is not the case. The remainder of the table, however, clarifies the matter. The more intelligent patrolmen leave the service rather early. It is quite possible that these patrolmen would have made better officers, but they did not remain long enough to be promoted. In another city the same tendency was found for the more intelli-

¹⁰ After Viteles.

gent to leave earlier, although the officers in this case made somewhat higher grades on the average than did the patrolmen.

Waitresses. A group of waitresses who had served 4 months to 15 years averaged 17 points in an intelligence test, their scores ranging from 4 to 33. At the same concern the waitresses who had served less than 4 months averaged 32 points in intelligence and their scores ranged from 15 to 45. Those of lower intelligence were manifestly more stable [24].

TABLE 33. AVERAGE INTELLIGENCE OF POLICEMEN¹¹

Lieutenants.....	58
Sergeants.....	55
Patrolmen (all).....	71
Patrolmen in service less than 9 months.....	72
Patrolmen in service 10 to 19 months.....	64
Patrolmen in service over 20 months.....	51

Salesmen. With retail clerks, the correlation of intelligence and length of service for one group of employees was $-.31$ and for another group $-.11$ [15, 266]. This gives a slight indication that the less intelligent ones tend to remain longer in the employ. With house-to-house salesmen no correlation was found. For routine salesmen a coefficient of $-.44$ was obtained in one company, and $-.46$ in another. In one of these groups of retail salesmen, only 30 per cent of those scoring over 70 points remained with the company $2\frac{1}{2}$ years, while 64 per cent of those scoring below 70 remained for at least that length of time. Apparently the routine nature of the work, its easy mastery, and its lack of an attractive future produced instability among the more intellectual men. For heating equipment salesmen similarly, the correlation was $-.26$. For life insurance salesmen, on the other hand, there was a small positive correlation ($.23$) between intelligence and length of service. The same thing was found with office specialty salesmen. In one company the correlation for sales managers was $.61$ and for experienced salesmen $.21$, while in another it was $.12$ for the managers and $.50$ for the salesmen. For the inactive salesmen in one company there was, however, a negative correlation of $-.42$ between stability and in-

¹¹ After Thurstone.

telligence. In general, with the lower grades of selling there is a slight inverse relation between intelligence and stability, while with the higher grades there is a slight positive relation.

Dissatisfaction and Intelligence. A bit of additional evidence as to the undesirability of too much intelligence in certain lines of work is obtained from a consideration of the attitude of different groups of workers and their varying degrees of satisfaction with their work. In a concern where considerable dissatisfaction was found, it was analyzed with reference to the status of the most dissatisfied employees. Test results were not available, but school retardation as manifested by age and grade at leaving school was noted as an indirect indication of intelligence. In the tool department where the work was fairly complex, the greatest dissatisfaction was found among the workers who were presumably the most retarded intellectually. In the inspection department, on the other hand, where the work was repetitive and monotonous, the retarded individuals showed the least dissatisfaction. The brighter persons were apparently happier in the more complex work and the duller individuals in the simpler work [22].

In a school for unemployed young persons a number of girls were stitching on a wide-meshed canvas of standard size and shape. The most intelligent ones experienced the greatest boredom and their output was the most variable. They could reach a high output, but would not maintain it. A girl of medium intelligence was the most effective worker; she liked the work. A girl of low intelligence improved enormously in her work but was disturbed by conversation [7].

Upper Critical Score. Considerations such as these have in some instances led to the use of an upper as well as a lower critical score. With the group of routine salesmen previously mentioned, a critical score of 70 was established [15, 262]. Scores *above* this were considered unfavorable. This was one of the cases of a negative relation between intelligence and selling. Only 37 per cent of the above-average salesmen scored over 70 points, while 62 per cent of the below-average salesmen exceeded this intelligence score. On the other hand, 63 per cent of the above-average salesmen scored less than 70, while only 37 per cent of the below-average salesmen fell below this critical score. In various other instances where the curve for stability takes the

shape of the upper one in Fig. 7, it has proved advisable to set a critical score at each end of the intelligence scale.

A person may be too bright for a given job just as he may be too dull. Such an individual quickly masters the job, reaches its limits, and becomes dissatisfied. His work may be very effective almost from the start, but he "burns out" and leaves the organization. The employer's desire to hire the most intelligent employee is sometimes due to the fact that inadequate training or supervision is given. An employee of high intelligence may be able to shift for himself more effectively at the outset, but he may not be so permanent an asset as the person a little lower in the intellectual scale. Vocational placement, then, does not involve merely the selection of the ablest man for the job as far as intelligence is concerned. Overstocking a low-grade job with high-grade personnel will tend to increase turnover. In evaluating intelligence with reference to vocational aptitude, not merely maximum intelligence but rather optimum intelligence should be considered.

The foregoing discussion should be qualified, depending on the general status of unemployment. When jobs are scarce an upper critical score may not be necessary. Under these circumstances, people who are actually too good for the job may nevertheless take it and do it satisfactorily. Whereas ordinarily they would be dissatisfied because of its routine character, if no other employment is in sight they may become reasonably well adjusted to the work emotionally. When jobs pick up, however, an upper critical score may be necessary again [23].

SUMMARY

There are occupations which require for their effective performance no specialized capacity, but rather general ability or intelligence. A consideration of the average intelligence of various groups of workers reveals an occupational hierarchy. The unskilled laborers are inferior in intelligence to the semi-skilled or skilled workers. These in turn are surpassed by persons in technical, business, or clerical work. Members of the professions come at the top of the scale. The theory is that an individual will in the long run attain about as high an occupational level in the hierarchy as his intelligence warrants. Hence these group

averages are tantamount to the intellectual requirements of the occupations in question. It is thus possible to locate an individual applicant at somewhere near the occupational level for which he is best fitted; with applicants with extremely high or low intelligence the assignment to occupations at the opposite extreme is manifestly inadvisable. Similar hierarchies are found for the various jobs within a single organization and for different types of salespeople. Retail clerks have the lowest average intelligence scores. They are surpassed by the wholesale and routine salesmen. These in turn are exceeded by real estate, insurance, and specialty groups, with salesmen for technical products and sales engineers at the top of the hierarchy. Apart from the other requisites of salesmanship, certain aspects of the occupation are more exacting in their intelligence requirements than are others. There are even some cases of adequate vocational adjustment for the higher grades of feeble-minded.

Intelligence tests are sometimes useful in surveying an organization or group of organizations. Such a survey throws light on the results attained by present employment methods and often raises further problems for analysis. For instance, one company found that the male office employees possessed higher average intelligence than the female. Different concerns in the same community were employing clerical workers of distinctly different intelligence levels. Several similar companies were attracting the same grade of applicants, but had marked differences in the resulting personnel. These findings pointed to the need for analysis of employment methods and policies.

Intelligence tests, like capacity tests, may be correlated with a criterion. Fairly high correlations have been found for clerical workers, office boys, operators in clothcraft shops, and for certain types of executives. The results with salesmen were more equivocal. There were indications of small negative correlations with intelligence for the lower grades of selling ability and small positive correlations for the higher grades. In some other occupations no correlation whatever has been found, and in a few instances of rather closely supervised work there have been appreciable negative correlations.

Critical scores for intelligence may be set in the same fashion as critical scores for special capacity tests. Some concerns main-

tain a set of critical scores for different jobs in their organization, especially office jobs. Workers falling below these critical points are not hired unless they possess some compensating qualifications.

In occupations in which there is a correlation between proficiency and intelligence, it is not necessarily desirable to employ persons with the maximum possible intelligence. Such individuals may learn readily and become effective workers soon after their induction, but in many instances it has been demonstrated that they do not remain long in the employ. With various types of office workers, cashiers, policemen, waitresses, and some of the lower grades of salesmen, more instability or turnover has been found among those of high intelligence than among those of average intelligence. While persons of very low intelligence may not have sufficient ability to learn effectively and perform their duties, those of very high intelligence may be too good for the job. It is not sufficiently exacting to hold their interest, their intellectual ability has insufficient outlet, and they become dissatisfied. This points in some instances to the necessity for an upper critical score. Applicants scoring above this amount are considered unsuitable material from the standpoint of permanency. Where intelligence is related to vocational aptitude, it is often desirable to consider not maximum intelligence but optimum intelligence.

REFERENCES

1. Abel, T. M. Study of a Group of Subnormal Girls Successfully Adjusted in Industry and the Community. *American Journal of Mental Deficiency*, 1940, 45, 66-72.
2. Beckham, A. S. Minimum Intelligence Levels for Several Occupations. *Personnel Journal*, 1930, 9, 309-313.
3. Bernstein, C. How Small Town Industry Makes Use of the Moron. *Industrial Psychology*, 1927, 2, 305-309.
4. Bills, M. A. Relation of Mental Alertness Test Score to Position and Permanency in Company. *Journal of Applied Psychology*, 1923, 7, 154-156.
5. Bingham, W. V. *Aptitudes and Aptitude Testing*. New York, Harper, 1937, 390 pp.
6. Bingham, W. V., and Davis, W. T. Intelligence Test Scores and Business Success. *Journal of Applied Psychology*, 1924, 8, 1-22.

7. Burnet, I. An Experimental Investigation of Repetitive Work. *Journal of the National Institute of Industrial Psychology*, 1924, 2, 18-23.
8. Burr, E. T. Adapting the Feeble-minded to Industry. *Industrial Psychology*, 1927, 2, 132-138.
9. Burr, E. T. Minimum Intellectual Levels of Accomplishment in Industry. *Journal of Personnel Research*, 1924, 3, 207-212.
10. Cowdery, K. M. Measures of General Intelligence as Indices of Success in Trade Learning. *Journal of Applied Psychology*, 1922, 6, 311-330.
11. Hartson, L. D. Intelligence and Scholarship of Occupational Groups. *Personnel Journal*, 1928, 7, 281-285.
12. Hildreth, G. Occupational Status and Intelligence. *Personnel Journal*, 1934, 13, 153-157.
13. Johnson, T. R. Reducing Deliverymen's Mistakes. *Industrial Psychology*, 1927, 2, 8-11.
14. Kenagy, H. G. Note in *Industrial Psychology*, 1928, 3, 128.
15. Kenagy, H. G., and Yoakum, C. E. *The Selection and Training of Salesmen*. New York, McGraw-Hill, 1925, 380 pp.
16. Kornhauser, A. W. Some Business Applications of a Mental Alertness Test. *Journal of Personnel Research*, 1922, 1, 103-121.
17. Miner, J. B. Standardizing Tests for Vocational Guidance. *School and Society*, 1921, 13, 629-633.
18. National Academy of Sciences. Psychological Examining in the United States Army. *Memoirs*, 1921, 15, 890 pp.
19. Otis, A. S. The Selection of Mill Workers by Mental Tests. *Journal of Applied Psychology*, 1920, 4, 339-341.
20. Pond, M., and Bills, M. A. Intelligence and Clerical Jobs. *Personnel Journal*, 1933, 12, 41-56.
21. Scott, W. D., and Clothier, R. C. *Personnel Management*. Chicago, Shaw, 1923, 643 pp.
22. Snow, A. J. Labor Turnover and Mental Alertness Test Scores. *Journal of Applied Psychology*, 1923, 7, 285-290.
23. Snow, A. J. Intelligence and Labor Turnover in the Taxicab Industry. *Industrial Psychology*, 1927, 2, 191-195.
24. Starch, D. The Use and Limitations of Psychological Tests. *Harvard Business Review*, 1922, 1, 71-80.
25. Thurstone, L. L. The Intelligence of Policemen. *Journal of Personnel Research*, 1922, 1, 64-74.
26. Viteles, M. S. Selecting Cashiers and Predicting Length of Service. *Journal of Personnel Research*, 1924, 2, 467-473.

Chapter XI

INTERESTS IN EMPLOYMENT PSYCHOLOGY

Occupational success depends on many things in addition to innate ability. Any employment man will immediately recall instances in which an applicant with the requisite ability was an occupational failure because he did not use that ability. The personnel psychologist will frequently predict an applicant's success on the basis of test score and then the man will fail to come up to expectations. This failure to exert himself may have been due to his lack of incentive or lack of interest. The first of these is not an employment problem. The difficulty does not arise until after the man is hired and it involves the consideration of methods of instruction, working conditions, wages, and various other incentives which motivate the worker. The problem of interest, however, is germane to the present discussion. Many a man is physically present at his work, but mentally absent. This is undesirable, for he is less apt to use his capacities effectively and is more apt to be discontented. To be sure, it is sometimes possible to modify an interest or to arouse one where it has not existed previously, but many applicants approach the prospective employer with pretty firmly established interests. Whether these are innate or acquired is of minor importance compared with their firmly fixed character. They give the worker a certain bias which may or may not be favorable to his success. Consequently the study of interests is a logical aspect of the employment program.

It is rather obvious that wide differences in interests exist between individuals. A casual consideration of one's acquaintances will reveal this. Some persons enjoy tinkering with tools or machinery, while others dislike to drive a nail. Some enjoy meeting people and talking with them, while others are content

with their own company. Some enjoy classical music, while others prefer jazz. Some are enthusiastic about art or literature, while others give it little attention. Some are scientifically curious about the reason for the things in their environment, while others are content to take things unquestioningly as they find them. Some of these individual differences in interest may be of vocational significance. It only remains to devise more effective means of ascertaining their existence and of evaluating their practical importance once they have been discovered.

Interest and Opportunity. The psychologist concerned with vocational adjustment is confronted with a serious discrepancy between vocational interests and existing vocational opportunity. This discrepancy does not hit the employment man as seriously as the vocational counselor but he should be cognizant of it. Applicants who come to an employment office for a job may be doing so as a second or third choice, thus making it more important for the employment man to "sell" them the job.

This discrepancy may be noted in a tabulation of a sample of young people who were seeking vocational advice and expressed their interest in one of four large categories [12, 81]. The frequency of their choices was compared with the actual frequency of opportunities of the same sort in the community where they were seeking a vocational adjustment. The data appear in Table 34. It is apparent, for instance, that none of the youngsters express an interest in unskilled laboring jobs, although 12 per cent of the available jobs in the community are of that

TABLE 34. DISCREPANCY BETWEEN INTEREST AND OPPORTUNITY
FOR YOUNG PEOPLE¹

	Expressed Interest	Opportunity
Unskilled labor.....	0	12
Skilled labor.....	10	27
Business.....	29	45
Professions.....	61	16
	100%	100%

¹ Laird (after Madsen).

type. On the other hand, 61 per cent of them seek a professional objective, whereas only 16 per cent of the opportunities lie in professional fields. Part of this discrepancy is due to the greater social value attached to a white-collar job—some workers are more concerned with social status than with the work itself. This notion does lead, however, to frequent maladjustment. Consequently when hiring a man for a skilled laboring job, it may be well to stress its greater opportunities, especially if the applicant expresses an interest in the white-collar type.

PERMANENCE OF INTEREST

Many workers come to the employment office with interests that are apparently rather firmly fixed. This question of permanence of interest has been studied statistically [22]. About 350 individuals were requested to estimate in retrospect their interests in certain school subjects—mathematics, history, literature, science, music, drawing, and manual work. They estimated their relative interest in these subjects in grade school, then in high school, and finally in college. While errors of recollection doubtless enter into such estimates, the results were sufficiently striking to carry a presumption of some permanence of the interests. Correlations of from .60 to .70 were found between interest at the age of 10-14 and at the age of 21. These interests, of course, involved only academic subjects, but some of them—for instance, the manual interest—might be of vocational significance.

A group of college women before graduation expressed their vocational interests, indicating in a list of vocations five choices in order of preference. Two years later they were sent a questionnaire asking for a similar record of vocational preference. Many of them had meanwhile been employed, and this might have altered their initial interest. However, 75 per cent of them still maintained the same vocation as their first choice, although 41 per cent changed their second choice.

A study of high school seniors, on the other hand, indicated that about half of them had changed their vocational intention at least once prior to the time of the study [6]. Even in a supposedly stable group, such as appears in *Who's Who*, 16 per cent changed their vocations at some time, presumably in many cases because of a shift of interest.

There are some indications that the permanence of vocational interest depends on the vocation [10]. College women were given an interest questionnaire twice, a year apart, and the correlations varied from $-.11$ for nurse to $+.86$ for physician. The correlations averaged about $.50$, indicating a fair degree of permanence, but this was obviously dependent somewhat upon the occupation.

If the vocational interest holds for a number of years the probability increases that it will hold subsequently [8]. One hundred students at the University of Kansas were followed up ten years after graduation. Many of them were in a vocation which had been their choice at an earlier period. Seventy-eight per cent of them had actually made that choice before high school, 72 per cent during high school, and 59 per cent during college.

In the studies discussed thus far, the technique involved simply asking the subjects to express interest in a series of vocations. The problem has also been investigated by the use of Strong's Interest Inventory, which will be described more fully below (p. 327). In this blank the subject checks a large number of items as to whether he likes or dislikes them; scoring patterns have been worked out based on people in various occupational groups. If the subject's questionnaire is scored in the usual fashion and his interests as indicated by these scoring patterns are simply ranked—for example, engineer first choice, chemist second—and if the same procedure is repeated five years later, the two series of ranks correlate on the average to the extent of $.75$ [21]. These data give a higher correlation than that usually found when dealing with a specific statement of vocational interests. The ordinary vocational interests may be subject to numerous extraneous factors such as the influence of friends or geographic location, whereas Strong's blank gets at the basic patterns even in fields which the subject has not considered specifically. He might, for instance, be interested in the same kind of things in which insurance salesmen are interested but never have thought of that particular vocational objective. These patterns of interest apparently are laid down fairly early and in advance of any specific vocational choice so they might be

expected to be more permanent. This suggests the superiority of a test like this as a measure of permanent interests.

There is thus sufficient indication of permanence of interest to make it worth the consideration of the employment psychologist. It seems at least characteristic of the majority of individuals, although in many cases there may be a shift. The layman is perhaps inclined to overestimate this permanence. A parent is going entirely too far in assuming that because the child plays with a toy train his destiny lies in a locomotive cab, or that his predilection for filling bottles with water presages an adult interest in pharmacy. Almost every boy at some time looks forward to becoming a policeman, a fireman, or a bandit. The employment man, on the other hand, may be inclined to underestimate the permanence of interest and to hire men on the basis of ability, disregarding interest entirely. This is probably unwise because of the demonstrated stability of interest and because of the relation of interest to satisfaction with one's work.

This does not necessarily mean that the interest is inborn. We probably do have an innate interest in loud sounds, bright lights, and moving objects. Our interest in mechanical rather than literary pursuits is doubtless influenced by our experiences in childhood or later. An interest in chemistry or physics reflects to a still greater extent the environmental factor. The practical point is, however, that if a man approaches a job with a definite interest pro or con, the safest procedure is to assume that the interest will persist, and it should therefore be reckoned with in occupational prognosis.

INTEREST AND ABILITY

There has been some discussion as to the relation between interest and ability [23]. In one instance a group of students arranged the courses in their curriculum (mathematics, history, literature, science, music, drama, and hand work) in order of their interest and subsequently ranked these same subjects according to what they considered their own ability therein. The correlations between rank for interest and rank for estimated ability averaged .89. The results were not so striking in another group of students when ranking for interest in college subjects was correlated with actual marks. It is rather probable, however,

that accidental factors were involved in the academic grades and that the individuals' estimates of their own ability came nearer to the real truth than did the grades obtained. This is further substantiated by the fact that estimates of ability and actual grades correlated to the extent of only .47. At any rate, there seems to be enough relation between interest and ability to be of some significance.

A group of men in the employment department of a YMCA expressed their vocational interests. They were also given an intelligence test. The preferred vocations were located in the occupational hierarchy discussed in the preceding chapter, and the intelligence required by the job in which interest was expressed was correlated with the actual intelligence of the man expressing that interest. The coefficient was only .38. When fairly liberal allowance was made, 36 per cent possessed more intelligence than that required for the job in which they expressed interest, while 15 per cent had less than the requisite intelligence. The conclusion is drawn that the correlation between interest and ability is not over .50.

A repetition of this procedure at the University of Utah and New York University yields a correlation between interest and estimated ability of about .60 on the average, with a range from .50 to .70. On the other hand, the correlations between expressed interest and actual ability as indicated by academic grades range from .04 to .57, with an average of about .30 [9].

In a study of design engineers and sales engineers various special engineering aptitude tests were employed. Some interest questionnaires were also given. The interests correlated with the special capacity tests to the extent of .50 [14].

The relation between ability and interest, then, is apparently not an extremely close one and there are, of course, obvious cases of lack of correspondence. A person may want to sing, but he may have a poor vocal apparatus; he may aspire to a berth on the police force, although he weighs only 110 pounds. Nevertheless, the studies just cited indicate a relation that is sufficiently close to merit some attention. It may be that one likes what he can do well. Or it may be that one devotes effort to the thing that he likes. In either case interests are worth considering from the employment standpoint. If the former alternative is true, interest

in a certain field would seem to indicate that the person had been successful in that general area and hence the interest might be diagnostic of probable success in related fields. If the latter alternative is true, it indicates the desirability of employing an individual for work in which he has some interest because he will then devote greater effort to it and use whatever ability he possesses.

METHODS OF MEASURING INTEREST

Questionnaire. Granted, then, that interests are of some importance to the employment psychologist, the question arises of how information regarding them may best be obtained. There are three general methods of approach to the problem of measuring interests: (1) by questionnaire; (2) by information tests; and (3) by more indirect methods. The most direct approach of course is to ask a person if he is interested in a certain occupation or to give him a list of occupations which he ranks in order of interest. This procedure has little to recommend it. Perhaps he has had no adequate contact with certain occupations so that he has no basis for judgment. Or he may think that he dislikes any kind of carpentry work because he has worked only at some poorly equipped, poorly lighted bench in an attic. There is also a possibility that he will not admit an actual interest because of the social status of the job.

The questionnaire is a more extensive though indirect procedure. The applicant may be questioned regarding previous vocational activities with a view to throwing light on his subsequent vocational interest. The following questions are typical:

1. Have you ever worked as a clerk in a store?.....
2. Have you ever conducted a house-to-house canvass?.....
3. What is the most responsible position you ever held?.....
4. What job that you have ever held did you like best?.....
5. Estimate how many hours during the past year you have spent working with tools, machinery, engines, and electrical apparatus
.....
6. Have you ever constructed a piece of furniture or household appliance?.....
7. Did you construct it because you wanted the appliance or because you enjoyed making it?.....

8. Do you think you could find out what was wrong with a clock that would not run?.....; an electric motor?.....; an automobile engine?.....
9. Have you ever written a story that appeared in print?.....
10. Have you ever taught or tutored anyone in a school subject?.....

Questions like these may be devised to cover a wide range of possible vocational interests. The selection of questions, of course, depends on the occupation for which they are to be used. If one is especially concerned with locating people who have been inclined toward social or mechanical vocations, the questions can concentrate particularly on these points.

One questionnaire designed to bring out interests in the field of engineering dealt with previous attitude toward specific situations involving mechanical items [16]. The first few items of the blank which were to be checked if they characterized the person were as follows:

1. Never care to use tools or handle machinery and avoid them
2. Use tools, make home repairs and maintain machinery *only* as it is necessary.....
3. Like to repair and overhaul engines, machinery, and all things around home.....
4. Like to build wagons, mechanical devices, electrical apparatus and do build some.....
5. Enjoy building machinery, engines, electrical apparatus, and instruments and do so continually.....

The questionnaire may also be devoted exclusively to avocational interests. Consideration of these may throw light on tendencies that will be of later vocational significance. The following questions are typical:

1. What are your principal hobbies?.....
2. What is your customary recreation in the evening?.....
3. What sports or games do you like to watch?.....
4. What magazines do you read regularly?.....
5. What books that you read within the last year interested you most?
6. Estimate how many hours during the past year you have spent in each of the following: driving an automobile.....; riding a

- motorcycle.....; horseback riding.....; hunting or rifle practice.....; swimming.....; tennis.....; golf.....; handball.....; other athletic sports.....
7. Assuming equal acting ability in each, which of the following do you prefer: dramas?.....; musical comedy?.....; vaudeville?.....
 8. In which of the following activities have you ever taken part: dramatics?.....; musical organizations?.....; debating?.....; politics?.....; public speaking?.....; reporting on a paper?.....
 9. Have you ever made a collection of: stamps?.....; coins?.....; postal cards?..... What else have you collected?.....
 10. In listening to radio what do you prefer: lectures?.....; concerts?.....; dance music?.....; news items?.....; logging stations?.....

Questions such as these may be made up to cover a great number of possible avocational interests which may be of vocational significance. The type of recreation one pursues may indicate his predilection for outdoor vs. indoor activity. His hobbies may give some clue to his inclination toward the mechanical. Miscellaneous activities will show something regarding propensity for literary, forensic, or physical work. For a particular occupation it may prove possible to determine the avocational interests that are of the greatest significance and devise questions to bring them out specifically.

There are some occupations that manifestly need an individual with social inclinations. Some of the questionnaire items may involve matters that will serve to indicate the social type of person. A few typical questions follow.

1. At what age did you learn to dance?.....
2. Estimate how many smokers, lodge meetings, card parties, and other social affairs for your own sex you have attended during the past year.....
3. Estimate how many mixed social affairs you have attended during the past year (include dance parties, socials, etc.).....
4. To what social clubs, fraternities, or business organizations do you belong?.....
5. What offices have you held in these organizations?.....

6. When single, did you prefer to room alone or with a roommate?.....
7. Do you enjoy going to the theater by yourself?.....
8. How frequently do you play solitaire?.....
9. Have you a very few close friends?.....a great many ordinary friends?.....or both?.....
10. With how many persons do you maintain a social correspondence?.....

Questions like these may be designed to bring out whether a person seems to enjoy the company of other people and to be more or less dependent on it or whether he is frequently satisfied to remain alone without social contact.

Strong's Interest Inventory. The techniques just described have their obvious shortcomings. One in particular is the danger that the subject may mark the items not truthfully as they characterize him but rather in a manner that he thinks will secure the job. It is possible to obviate this difficulty to some extent by having the test composed of so many items that he cannot figure out exactly how to mark them in order to make a favorable impression. This procedure has formed the basis of a number of interest inventories. These consist essentially of a considerable series of items which the subject checks according to whether he likes or dislikes them. This procedure was initiated originally at Carnegie Institute of Technology by Ream and Freyd, later developed somewhat by Cowdery, and finally standardized by Strong in a form that is widely used.

The blank itself comprises something over 400 items which the subject marks L, D, or I, according to whether he likes, dislikes, or is indifferent to them. The first section of the blank deals with occupations such as actor, advertiser, architect, Army officer, artist. The next section deals with amusements such as golf, fishing, hunting, driving an automobile, taking long walks, playing checkers. The third portion involves school subjects such as algebra, agriculture, arithmetic, art, Bible study, botany. The fourth deals with activities such as repairing a clock, adjusting a carburetor, handling horses, raising flowers and vegetables, giving first aid, making a radio set, debating. The fifth portion involves peculiarities of people such as progressive people, conservative people, energetic people, absent-minded people, people

who borrow things, optimists, pessimists. Another section of the test lists ten activities and the subject selects the three which he would most enjoy doing and the three which he would enjoy least. He also has to make a choice between various pairs of items such as streetcar motorman, streetcar conductor; policeman, fireman; persuade others, order others; definite salary, commission on what he does.

Scoring patterns are worked out for various occupations by essentially the following technique: Suppose that the test is given to several hundred engineers and several hundred persons who are not engineers. If we assume now that the engineers on the whole are in a job which interests them we may try to derive a scheme for scoring the test which will differentiate engineers from non-engineers. We analyze the data item by item. For example, if on the first item, actor, 70 per cent of the engineers say they like it and only 30 per cent of the non-engineers do so, this looks like a differential item. If, however, 70 per cent of one group and 68 per cent of the other group like it, it does not serve to differentiate the two groups and should receive little weight in a scoring pattern. In actual practice a fourfold table is constructed; in the first row are the persons in the occupation, in the second row the persons not in it, and in the two columns a breakdown according to likes and dislikes for the item in question.

A statistical question arises at this point, namely, whether the difference between the two percentages is significant. In the above example undoubtedly the difference between 70 per cent and 30 per cent represents a real difference. It is very improbable, if we took another sample of engineers and non-engineers, that the tendency would be reversed, that is, that more of the non-engineers would express a liking for the item. As the percentages come closer together, the chances increase that we might actually get the opposite result on repetition with another sample. It is also true that the larger the number of individuals included in our sample the more probable it is that we have a typical group and that the result would not be greatly changed on repetition.

It is possible to evaluate this point statistically and to determine whether the difference between the percentages actually

is significant in the above sense. This involves computing the standard deviation of the difference. The formula is:

$$\sqrt{\frac{p_1q_1}{N_1} + \frac{p_2q_2}{N_2}}$$

where p_1 indicates the percentage of engineers liking the item, q_1 the percentage of the engineers who do not express a liking for the item, N_1 the number of engineers involved, p_2 the percentage of non-engineers liking the item, q_2 the percentage of non-engineers who do not express a liking for the item, and N_2 the number of non-engineers. If the actual difference between the percentages is twice this standard deviation of the difference, we know from the theory of probability that the chances are about 98 out of 100 that if we repeated the experiment with other samples the differences would still be in the same direction. If the difference is three times its standard deviation, it is almost certain that the difference would not be reversed on repetition and we would consider the difference as "statistically significant." Few statisticians would attach much significance to a difference less than twice its standard deviation and it is common practice to demand a ratio of 3.0.

In the standardization of Strong's Interest Inventory, items in which the two groups differ significantly in the above sense are retained in the scoring pattern for the occupation in question. By a further extension of this principle, appropriate weights are derived for the various items in accordance with their differential importance. These weights range from 2 to 27. Their computation may be facilitated by specially prepared tables [18, 255]. Strong has worked out patterns for scoring the blanks for interest in some 40 different occupations. The subject marks the blank only once, but it can be scored separately for each occupation in question. This is a time-consuming procedure, but by punching the results into Hollerith cards and having appropriately prepared master cards it is possible to expedite the scoring.

Strong's Interest Inventory has been more widely used than has any other device of this sort for evaluating interest. It is well to mention the application of factor analysis (cf. p. 107) to some of this material [24]. The various patterns are intercorrelated at the outset; for example, we determine the extent to

which those who have a high interest score for engineering score similarly high for salesmanship. Starting with this array of intercorrelations, by appropriate statistical techniques we attempt to discover how many factors would be necessary to account for all these intercorrelations and we also determine the loading for each profession or occupation in each factor. Then from an inspection of the loadings we speculate as to the nature of the factors. An analysis made by this technique with intercorrelations between 18 occupations yielded 4 factors. The factor loadings are given in Table 35.

TABLE 35. FACTOR LOADINGS FOR OCCUPATIONAL INTEREST PATTERNS¹

	I	II	III	IV
Advertising	-.48	.66	-.21	.22
Art45	.70	-.18	-.31
C.P.A.	-.04	.32	.00	.56
Chemistry98	-.21	-.15	.06
Engineering84	-.36	-.22	.16
Law	-.23	.77	-.12	.44
Ministry09	.51	.62	-.30
Psychology77	.47	-.04	-.28
Teaching36	.15	.68	-.22
Life insurance	-.82	-.02	.27	.45
Architect83	.26	.16	.05
YMCA secretary	-.23	.00	.90	-.37
Farming71	-.54	.01	.18
Purchasing agent	-.05	-.79	.01	.44
Journalism	-.15	.84	-.28	.25
Personnel	-.30	-.26	.66	-.19
Real estate	-.76	-.07	-.06	.58
Medicine71	.33	-.26	-.09

Looking at the first factor loadings in the first column, we note high loadings for chemistry, engineering, psychology, architect, and farming; very low ones for public accountant, the ministry, and purchasing agent; and negative loadings for advertising, life insurance, and real estate salesmen. It appears that this first factor might be an interest in science. For the second factor we have a fairly high loading for advertising, art, law, ministry,

² After Thurstone.

psychology and journalism. This factor is tentatively identified as interest in language. In the third column the high loadings are for ministry, teaching, YMCA secretary, and personnel work, which suggests an interest in people. The loadings usually decrease as the number of factors increases, but still the largest ones are of interest. In the fourth column we note heaviest loadings for public accountant, law, life insurance, purchasing agent, and real estate salesman, which suggests interest in business as a factor. Thus it appears that the interest patterns involved in Strong's blank are made up of combinations of these four basic factors. Advertising interest, for example, is made up especially of interest in language and in business, with an absence of interest in science and people. Chemistry interest is made up almost entirely of scientific interest. Law involves interest in language and in business, and so it goes.

A further analysis along the same line has been reported by Strong [20], using a somewhat larger number of occupational ability patterns and coming to essentially the same result with, however, the possibility of a fifth factor which could not be so clearly identified.

One other analysis along this line may be mentioned [7]. A sample of pre-medical students and the 19 scoring patterns for that many occupations were used, and four factors were derived with their appropriate loadings. Further interest, however, lay in the possibility that certain of the occupations might be "fundamental" in the sense that if a person's score in them was known, his score in any other occupation could be easily computed by weighting those few as in a regression equation (p. 233). Analysis did reveal that four occupations—namely, physicist, journalist, minister, and life insurance salesman—were fundamental in the above sense. Then, with any other occupation as the criterion, regression equations were worked out using these fundamental occupations as the independent variables; that is, they corresponded to the "tests" in the usual procedure of standardizing aptitude tests against a criterion. After the weights were derived for a given occupation its multiple correlations on the whole were .80 or better, with the exception of two—farmer and accountant. This procedure may somewhat facilitate scoring the blank when punched card machines are not available. The subject's responses

are weighted in detail to obtain the four fundamental scores, and a table of weights of these four scores is consulted for any given occupation; it is then necessary merely to add the weighted scores and have the interest score for that occupation.

Information Test. Instead of relying on the subject's own statements regarding his interests or preferences which in some cases may be influenced either by the use which he thinks is to be made of them or by his efforts to make such answers as will give favorable consideration to his case, it is possible to approach the matter more indirectly but more objectively. There is some ground for the assumption that if a person is interested in a certain field he will pick up information about it—be more familiar with the terminology and with less obvious details that would presumably be overlooked by a person who lacked that interest. Consequently, an information test may give some indication of interest if the items are carefully selected. It is insufficient to ask questions that anyone would be able to answer from casual observation. It is necessary to go further into details such as one would not encounter unless he had made a definite effort to pursue the particular line under consideration. Below are given a few items from an interest test for agricultural engineers [5]. In selecting items ordinary things that the students would meet in their everyday work in the college course were avoided. Technical journals were consulted and out-of-the-way facts were selected on the theory that the student who was interested in this profession would naturally go beyond the ordinary required work of the classroom and would read additional things like technical journals. In each item the subject checks the correct alternative.

1. SUNLIGHT is an OXIDIZING AGENT; HUMIDIFIER; POISON; DISINFECTANT; TOXIN.
2. A TRACK-LAYING TRACTOR is used for ROAD-ROLLING; PAINTING; MILKING; LAYING TRACKS; PLOWING.
3. A SURVEYOR'S LEVEL is used for GRADING ROADS; FINDING AREAS; DETERMINING DIFFERENCES IN ELEVATION; MEASURING ANGLES; MEASURING PERPENDICULARS.
4. A PINION is a LOCK-NUT; SMALL GEAR; WHEEL; KEY; RACK.
5. A CONVEYOR BELT is used for moving GRAIN; GASOLINE; MOLASSES; BARRELS OF SALT; LAMP CHIMNEYS.

6. DRAIN TILES are used for ELECTRICAL CONDUCTS; BUILDING GARAGES; LOWERING SOIL WATER TABLE; PAVING; ROOFING.
7. The BEST BLOWER BELTS are made of LEATHER; RUBBER; JUTE; SILK; HEMP.
8. CREOSOTE is a FUNGICIDE; VARNISH; CATALYZING AGENT; BREAD FLOUR; SUGAR CLARIFIER.
9. SOIL STACK is a term used in BOILER FITTING; SURVEYING; CEMENT MANUFACTURE; PLUMBING; SOIL ANALYSIS.
10. The BEST HEAT INSULATOR is WATER JACKETING; SHEET CORK; HARD RUBBER; POWDERED ROSIN; CEMENT BLOCKS.

To cite another instance of an information test used as an index of interests, a concern wished to know especially about the social interests of its applicants, whether they were "good mixers" and whether their interests had led them into a wide social experience. Information questions were devised to cover a considerable range of possible social interests [15]. A few items of each kind will be cited by way of illustration. The test involved items that were socially acceptable, that dealt with sports, and that were perhaps socially questionable. Some of the socially acceptable items are as follows. As in the preceding example the subject checks the correct alternative in each item.

1. Which of the following requires chairs? LONDON BRIDGE; FLYING DUTCHMAN; THREE DEEP; GOING TO JERUSALEM.
2. In what organization is 11 o'clock of special significance? ELKS; ODD FELLOWS; MASONS; KNIGHTS OF COLUMBUS.
3. In the song what follows the words "Blest be the tie that binds"? "US IN THY KINGDOM, LORD"; "MY FAITH ON CALVARY"; "LOVED ONES OF KINDRED MINDS"; "OUR HEARTS IN CHRISTIAN LOVE."
4. What is a caucus? A NATIONAL POLITICAL CONVENTION; AN OFFICIAL COUNTY ELECTION; A MEETING OF POLITICIANS WITHIN ONE PARTY; A SECRET POLITICAL MEETING IN VIOLATION OF THE LAW.
5. What is "French leave"? A DANCE; VERY FEW ODDS AND ENDS LEFT OVER; PERMISSION EASILY OBTAINED; SLIPPING AWAY WITHOUT NOTICE; SHOWING VERY POLITE MANNERS.

Some of the sports items are as follows:

1. What is the nickname of the Chicago Nationals? CARDINALS; BRAVES; WHITE SOX; CUBS.
2. Which of the following clubs has a wooden head? CLEEK; BRASSIE; NIBLICK; MASHIE.

3. What kind of a blow is a haymaker? HOOK; UPPERCUT; BROADSIDE; JAB.
4. What is the score when all 10 pins are knocked down? STRIKE; LITTLE SLAM; SPARE; BREAK.
5. What kind of a race is a derby? TROTting; PACING; RUNNING; HURDLING.

Some of the possibly questionable items are as follows:

1. How many spots on dice make a Little Joe? THREE; FOUR; SEVEN; ELEVEN.
2. Which kind of wine is the strongest? CLARET; CHAMPAGNE; SHERRY; BURGUNDY; BORDEAUX.
3. What beats a flush? FOURS; STRAIGHT; THREE OF A KIND; TWO PAIR.
4. What is the name applied to short lively chorus girls? KITTENS; PONIES; BABY DOLLS; FOOTLIGHT DODGERS.
5. Which of the following is best for jazz dancing? WALTZ; FOX TROT; PAUL JONES; MINUET.

There is a possible error in the information test as a measure of interest, at least when dealing with items that are socially questionable. As suggested earlier, the subject taking the test may become suspicious as to its purpose, and realize that correctly answering questions regarding poker will reveal the fact that he is familiar with the game. He may hesitate to commit himself for fear it will be held against him. This error is seldom involved when dealing with items about which no ethical question might be raised, but if some questionable ones are included they may create an unfavorable attitude toward the rest of the test. It is advisable to put such items at the end of the program so that if an atmosphere of suspicion is developed, it will not affect the results of any other tests or items.

Indirect Methods. Other methods of approaching interests are even more indirect. These methods are still in the experimental stage and too much stress should not be placed upon them until they are further validated. One of them involves what is ostensibly a memory test. The subject is given pairs of words and is required to associate the two words of each pair so that subsequently, when the first word of a pair is given, he can recall the second word that went with it. However, the pairs are chosen

according to two principles. Some of them form perfectly ordinary associations such as "dog—cat," "brook—river," whereas others, perhaps alternate pairs, involve associations dealing specifically with the type of work in question. It is probable that the person who is specially interested in the work will more readily associate the two words pertaining to it than will the person who is not. The following pairs of words are taken from such a test designed for agricultural engineers:

letter	stamp	brass	bearing
formula	equation	spider	spin
diamond	spade	level	transit
liquid	hydrometer	church	tower
watch	time	gasoline	kerosene
work	force	ocean	fish
rain	umbrella	engine	windmill

Alternate pairs, it is to be noted, deal with items familiar to agricultural engineers while the remaining pairs involve associations that are familiar to everyone. The presumption is that persons with agricultural engineering interests will more readily associate "formula" and "equation," "liquid" and "hydrometer," etc., than will people who do not possess such interests. If, however, the actual number of words recalled for the crucial pairs is taken as the final score, an error is introduced. One individual who has a profound interest in this profession may make a low score on such words, not because of lack of interest but because of poor memory, while another person with little interest but good memory may surpass him. This error may be obviated by taking the score on the crucial words relative to that on the normal words. The latter establishes the individual's general memory ability, and it is possible then to note by what percentage his performance on the crucial words exceeds or falls short of normal. If one individual does 10 per cent better on the crucial words than on the normal and another does 10 per cent worse on the crucial than he does on the normal, the former presumably has greater interest in the matter under consideration, regardless of the intrinsic memory ability of the two individuals.

A different approach has been made with a form of cancellation test. The subject is provided with a text containing irrele-

vant words which are to be crossed out. In some instances the material is of an ordinary uninteresting sort. In other cases it is designed to appeal to a particular interest. The following example is a portion of such a test designed to locate persons who are ambitious and particularly interested in success and achievement:

Part I

Advertising plays way today a very why conspicuous yet role in the yes management with of a business. It wan has assumed such proportions win in recent years war that it won is difficult to ton estimate the tan exact place tin which it occupies tip in commercial tub affairs. Over sib two-thirds of the son cost of maintaining a see newspaper or sun magazine is derived saw from advertising say space.

Part II

Suppose that gun it is success jot you want. There are few joys in this world that lab can compare lit with the joy of met achievement. Set your men mark and mat start climbing toward it. You mob will reach mud it if you keep mut at it. Be persistent pat and be pin patient. If you are in put Maine you can not wish rip yourself in rug California. But you'll sun get there sometime ton if you start tan and keep going tub even if you go rim on your hands tow and knees.

Part I, which is only a brief excerpt from the original test, is of an ordinary expository character with little appeal to any fundamental interest or tendency. Part II, however, is a "pep-talk" such as might appeal tremendously to a certain type of individual. Some persons read this kind of literature avidly and are much engrossed with the notion of personal success and "getting there." The theory of the test is that such people will become so wrapped up in the passage while going through it that they will overlook many of the irrelevant words which they are supposed to cross out.

In this test, as in the preceding, it is necessary to allow for the individual's intrinsic ability in this particular sort of task. One person may naturally be less efficient than another in detecting irrelevant words or in speed of reading and hence make a low score on Part II, not because of greater interest but because of lack of ability of this type. The uninteresting passage, however, serves as a control and indicates the individual's actual ability

in this kind of performance. The results for Part II may then be taken relative to this, provided identical time limits are used in the two cases. The presumption is that the lower the score made by a subject on the "pep-talk" relative to the normal text, the greater was his interest in the passage.

Similar tests may be developed in which instead of checking irrelevant words the subject corrects misspelled words or at least locates them. He may even have some simple task such as crossing out every "u" in the passages.

Other indirect procedures for evaluating interests bear passing mention although they are less applicable to the employment office. Among children much can be discovered by simply watching their play activity, both the type of games and the objects with which they play. Various objects can be presented deliberately and note made of the subject's choice—for instance, mechanical contrivances vs. books. With children, at least, information as to what books they have read or would like to read may be significant. If there is doubt about the veracity of their statements of what they have read, they may be given information tests covering a wide range of reading. Failure to identify some of the characters in a story would suggest that the subject had not read it in spite of his statement to the contrary.

An indirect measure of interest that has been employed in some investigations is the topic of spontaneous conversation. The technique involves listening to conversations in a large variety of situations, noting the topic and analyzing the results with reference to variables such as the sex of the persons participating in the conversation. It was found that men frequently talked about business or sports, whereas women discussed clothes and decoration or people [13]. Possibly these differences in topics of conversation did reflect actual basic sex differences in interests. This procedure might conceivably contribute to the solution of employment problems; for example, a member of the employment staff could listen to conversations in the waiting room, or a dictaphone might even be planted there. Such conversations, however, might not be typical because the persons in the waiting room would be talking about vocational problems more than they would normally and thus would not manifest their real interests.

EVALUATION OF MEASUREMENT OF INTERESTS

Various techniques for measurement of interests having been described, it is desirable now to indicate something regarding their validity. As suggested earlier, interests play a greater role in vocational guidance than in vocational selection. However, the personnel man is nevertheless concerned with the broad problem of vocational adjustment, and if his employees are in work in which they are interested they will probably be more efficient as well as happier.

Strong's Interest Inventory. Most of the experimental work on interests in recent years has centered around Strong's Interest Inventory. One way of evaluating the techniques is to take a given occupational pattern—for example, that for engineers—test persons in other occupations, and score their blanks by this same pattern. If lawyers and clergymen, for instance, make engineering interest scores comparable to those of the engineers themselves, the test cannot be considered unique in diagnosing engineering interest. A few studies of this sort will be presented.

It has become standard practice after developing the scoring pattern for a given occupation such as engineering to secure data for a sample of persons in that occupation and select the 75 per cent with the highest scores. The range of scores for this 75 per cent constitutes Class A in engineering interest. The range from this point down to the lowest score made by any of the engineers is designated Class B. Anything outside of this is Class C. Results of a study of engineers are summarized in Table 36 [19]. It gives in the first column of figures the percentage of the occupational group indicated who make engineering interest scores in Class A as described above. The next column gives the percentage of the occupational group falling in Class B; and the last column the percentage in Class C. In the first row 75 per cent of the engineers fall in Class A and 25 per cent in Class B by definition. However, in the remainder of the column for Class A it is to be noted that the percentages are very small. Physicians, personnel managers, teachers, and artists, when their blank is scored by the engineering interest pattern, fall far short of making the scores made by the engineers in that same pattern. This tends

TABLE 36. INTEREST INVENTORY PATTERN OF ENGINEERS IN COMPARISON WITH MEN IN OTHER OCCUPATIONS³

	Per Cent with Interest Scores of 75% of Engineers Scoring Highest (Class A)	Per Cent with Interest Scores of 25% of Engineers Scoring Lowest (Class B)	Per Cent with Interest Scores Outside of Engineers (Class C)
Engineers.....	75	25	0
Physicians.....	9	42	49
Personnel managers....	6	33	61
Teachers.....	8	29	63
Artists.....	4	29	67
Bankers.....	2	31	67
Office workers.....	5	25	70
Lawyers.....	5	25	70
C.P.A.....	2	25	73
Insurance salesmen....	0	29	71
Authors.....	4	18	78
Clergymen.....	0	10	90

to prove that the pattern actually does differentiate engineering interest from interest in other vocations.

A similar study for women may be cited in which teaching was used as the scoring pattern. The arrangement of the table is the same as in the preceding. Here again the percentages of the occupational groups other than teachers in Class A are small indeed. Thus the scoring pattern is quite differential [11].

Validation Against a Criterion. A few investigations have been made in which interests have been validated against a criterion of actual effectiveness on the job. The studies of the relation between interest and ability reported at the beginning of the chapter dealt primarily with academic ability, and there were indications of a fair correspondence. Similarly, some relationship might be expected between interest and success in a job on the general theory that people who are interested work harder, or the

³ After Strong.

TABLE 37. INTEREST INVENTORY PATTERN OF TEACHERS IN COMPARISON WITH WOMEN IN OTHER OCCUPATIONS⁴

	Per Cent with Interest Scores of 75% of Teachers Scoring Highest (Class A)	Per Cent with Interest Scores of 25% of Teachers Scoring Lowest (Class B)	Per Cent with Interest Scores Outside of Teachers (Class C)
Teachers.....	75	22	3
Housewives.....	9	43	48
Stenographers.....	2	47	51
Retail saleswomen.....	5	37	58
Emporium saleswomen..	5	23	72
Business women.....	5	22	73
Authors.....	0	23	77

greater effectiveness in the job itself brings a greater interest through feelings of accomplishment.

One such investigation dealt with casualty insurance salesmen. Among other tests they took the Strong Interest Inventory which incidentally proved to be the most valid of a number of other schedules which they filled out. Their blanks were scored by means of the patterns developed by Strong for life insurance salesmen and real estate salesmen. The following weighting scheme was arbitrarily adopted to convert letter grades to numerical scores:

A	3
B+	2
B	0
B-	-2
C	-3

The two scores (life insurance and real estate) for an individual were totaled and these totals compared with the criterion, which consisted of estimates by the managers. These latter involved three categories—outstanding success; success and fair plus; fail-

⁴ After Hogg.

ure and probable failure. They are compared with five categories of interest in Table 38. The entries in the table are simply the per-

TABLE 38. RELATION OF INTEREST INVENTORY SCORES TO SUCCESS IN SELLING CASUALTY INSURANCE⁵

	Interest Scores				
	+6	+4 to +5	+3 to -2	-3 to -5	-6
Outstanding success...	25	16	11	8	4
Success and fair plus...	53	56	47	39	20
Failure and probable failure.....	22	28	42	53	76

centage of those in the column who fall in a particular row. For example, of those who rated 6 points in the interest inventories, 25 per cent were "outstanding successes" in the estimates of their managers, while 53 per cent were rated as "success or fair plus." The total of each column is 100 per cent. Almost 600 men are represented in the table and it does indicate some validity for this interest test [2].

The foregoing procedure was repeated by the same investigator in 1940. The results were essentially the same, except that the proportions of success and failure appeared to vary with the different ages; therefore a further study was made with age groups separated. The percentage of failure, however, decreased with increased score on the interest inventory, with the exception of salesmen under the age of 24; for them a combined score of +5 or +4 was more advantageous than a +6. It was concluded that some of the men were unconsciously bluffing and that they overcompensated and therefore scored +6 on the sales items. There was a further suggestion that a secondary interest in some other occupation such as law might be favorable [3].

Another investigation of life insurance salesmen employed various personality tests, intelligence tests, and also an interest inventory. On the basis of the critical scores established for the personality and intelligence tests, about 69 per cent of the men

⁵ After Bills.

with the best sales records in getting new business would have been hired. By adding the Strong Interest Inventory to the battery this percentage was raised to 75. This indicates then that the interest test did add a little to the validity of the battery [17].

The Worker Analysis Section of the U.S. Employment Service, in developing a battery of tests for selecting department store salespersons, included the interest questionnaire which they had developed for subprofessional occupations, much along the line of the Strong Interest Inventory [18, 101].

In validating the Strong Interest Inventory or using it for purposes of guidance or selection one shortcoming is quite obvious, namely, the patterns that have been standardized and published are mostly at the professional and semi-professional level. In fact, about 82 per cent of the patterns are in these levels, which comprise only 14 per cent of the employed workers. An effort to derive other patterns or at least study the distribution of existing ones through the range of the population was made at Minnesota in connection with the Employment Stabilization Research Institute [1]. For one thing the investigators made up a standard sample of persons from various occupations in about the same proportion as those occupations existed in the city. For instance, if 5 per cent of the workers in the city were electricians, this proportion of the electricians would be included in the sample. When applying the scoring patterns for the basic interests in science, language, people, and business (p. 330), this particular group did not score high in science, language, or people, but the clerical pattern of interest was quite common in this group. The analysis was facilitated by using "standard scores" (p. 190) based on the standard sample. People in commercial occupations lacked the scientific interest patterns. Machinists showed some interest in science like the physicists and chemists. The better janitors showed interest in technical items. Laborers were not clearly differentiated from people in general. Obviously further investigation of subprofessional occupations is necessary before such interest inventories can be of maximum usefulness.

Information Test Results. In addition to evaluating questionnaires as a measure of interests, the information type of test has been evaluated. The one described above for social interests was designed originally for use with salesmen. The complete test

was given to a considerable number of salesmen who were divided into five groups on the basis of their ability. The information items were then tabulated to determine which were most differential of the groups. It proved possible in this way to select a fairly satisfactory set. A poor score on the test served rather definitely to indicate a poor salesman, although a high score did not always insure a good salesman. Apparently a lack of the interests that were covered in the test tended to render one poor salesmanship material, but their presence was not of itself sufficient because other factors not measured by this test were essential. A consideration of the most differential items reveals a tendency for the good salesman to be a man who has accepted social responsibility. One who is manifestly lacking in social experience seems to have poor chances for success in this line of work [15].

Results with Indirect Methods. The indirect methods of measuring interest have scarcely proceeded beyond the experimental stage. The method involving memory for word pairs described above was tried with a group of agricultural engineering students and the score on the test was correlated with an estimate by instructors as to "interest and industry." The test score—i.e., the memory for words related to agricultural engineering as compared to that for ordinary words—was found to correlate to the extent of .30 with estimated interest. The results were complicated by the fact that the test correlated with ability to an even greater extent, but there is some indication at least of possibilities in this method. The other indirect method above mentioned—the cancellation test in which it was supposed that interest in the passage would detract from efficiency in canceling irrelevant words—correlated $-.30$ with estimated interest. This negative correlation is also in conformity with the theory of the test. More work must necessarily be done with the indirect methods before any great validity is attached to them, but they are cited to show the possibility of measuring interest in these rather indirect ways [4].

SUMMARY

Occupational success depends on other things than ability alone, and interest is one of them. There is some indication that

interests are rather permanent, and hence it is necessary to reckon with them in the employment situation rather than to rely on their changing to meet conditions. There is also some relation between interest and ability. Whether the ability motivates the interest, or vice versa, has not been determined. In either instance, however, interests are to some extent diagnostic of what the person will ultimately do in the occupation.

Several methods have been used to determine systematically a person's interests. A questionnaire may be devised dealing with previous occupational, with avocational, or with social interests, any of which may be of practical significance. Instead of using questions to be answered, the procedure is sometimes varied by having the subject check a list of items according to whether he likes or dislikes them. By comparing the responses of people in different occupations item by item it is possible to derive scoring patterns for the interest inventory. Information tests are sometimes used as a measure of interest on the theory that a person who is interested in a certain field will go out of his way to obtain more information about it and will remain "set" for anything pertaining to it, so that he will in the long run be able to give a better account of himself in an information test involving items in this field. Still more indirect methods have been attempted. In what is ostensibly a memory test, in which some items appealing to a certain interest are mingled with other ordinary items, it is assumed that relatively more of the former will be retained by a person with that particular interest. In a test involving cancellation of irrelevant words in a text, it is assumed that if the content of the test appeals especially to the individual's interest he will become engrossed in it and mark relatively fewer of the irrelevant words.

These methods have been evaluated by administering the measurements or tests to certain occupational groups or to groups known to have some fundamental difference in interest and determining which items serve most clearly to differentiate the groups. It was possible from a list of items for which the subjects expressed their like or dislike to derive scoring patterns that are fairly unique for some 40 occupations. The information test as a measure of interest proved of some value in discriminating different degrees of success in selling. There were indications

that the successful salesman was an individual who had accepted social responsibilities. The more indirect methods of measuring interests have been tried to only a slight extent, but the correlations with estimated interest were somewhat encouraging. The whole matter of measuring interest and using such measurements in a practical way is still much in the experimental stage, but satisfactory progress is being made.

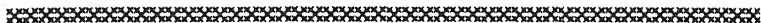
REFERENCES

1. Berman, I. R., Darley, J. G., and Patterson, D. G. Vocational Interest Scales. *Bulletin of Minnesota Employment Stabilization Research Institute*, 1934, 3, No. 5, 35 pp.
2. Bills, M. A. Relation of Interest Scores in Strong's Interest Analysis Blank to Success in Selling Casualty Insurance. *Journal of Applied Psychology*, 1938, 22, 97-104.
3. Bills, M. A. Selection of Casualty and Life Insurance Agents. *Journal of Applied Psychology*, 1941, 25, 6-10.
4. Burt, H. E. Measuring Interests Objectively. *School and Society*, 1923, 17, 444-448.
5. Burt, H. E., and Ives, F. Vocational Tests for Agricultural Engineers. *Journal of Applied Psychology*, 1923, 7, 178-187.
6. Douglas, A. A. Vocational Interests of High School Seniors. *School and Society*, 1922, 16, 79-84.
7. Dwyer, P. F. Analysis of 19 Occupational Interest Scores of Strong's Vocational Interest Test Given to 418 Students Entering the University of Michigan Medical School During the Years 1928, 29 and 30. *Journal of Applied Psychology*, 1938, 22, 8-16.
8. Dyer, D. T. Relations Between Vocational Interests of Men in College and Their Subsequent Occupational Histories for Ten Years. *Journal of Applied Psychology*, 1939, 23, 280-288.
9. Fryer, D. Interest and Ability in Educational Guidance. *Journal of Educational Research*, 1927, 16, 27-39.
10. Gaw, E. A. Occupational Interests of College Women. *Personnel Journal*, 1928, 7, 111-114.
11. Hogg, M. I. Occupational Interests of Women. *Personnel Journal*, 1928, 6, 331-337.
12. Laird, D. A. *How to Use Psychology in Business*. New York, McGraw-Hill, 1936, 378 pp.
13. Landis, M. H., and Burt, H. E. A Study of Conversations. *Journal of Comparative Psychology*, 1924, 4, 81-89.
14. Moore, B. V. Personnel Selection of Graduate Engineers. *Psychological Monographs*, 1921, 30, whole no. 138, 85 pp.

15. Ream, M. J. A Social Relations Test. *Journal of Applied Psychology*, 1922, 6, 69-73.
16. Rogers, H. S., and Holcomb, G. W. An Inventory of Engineering Motives. *Journal of Applied Psychology*, 1933, 17, 302-316.
17. Schultz, R. S. Test-selected Salesmen Are Successful. *Personnel Journal*, 1935, 14, 139-142.
18. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
19. Strong, E. K., Jr. Vocational Guidance of Engineers. *Industrial Psychology*, 1927, 2, 291-298.
20. Strong, E. K., Jr. Classification of Occupations by Interests. *Personnel Journal*, 1934, 12, 301-313.
21. Strong, E. K., Jr. Permanence of Vocational Interests. *Journal of Educational Psychology*, 1934, 25, 336-344.
22. Thorndike, E. L. Early Interests: Their Permanence and Relation to Ability. *School and Society*, 1917, 5, 178-179.
23. Thorndike, E. L. The Correlation Between Interests and Abilities in College Courses. *Psychological Review*, 1921, 28, 374-376.
24. Thurstone, L. L. A Multiple Factor Study of Vocational Interest. *Personnel Journal*, 1931, 10, 198-205.

Chapter XII

RATING SCALES



PURPOSE

Study of Non-measurable Traits. At the outset of the preceding chapter the point was made that capacity and ability do not constitute the whole story in predicting vocational aptitude. It is not merely a question of what the applicant *can* do, but of what he *will* do. He may be able to become a good calendar man, but in actual practice he does not try to make the most of his opportunity to learn to operate the machine and so never succeeds. A man may have the requisite intelligence, or memory, or speed of reaction for a given job, but he may lack industry, initiative, tact, enthusiasm, persistence, or other traits or attitudes or tendencies that are needed to supplement his ability in order to make him a successful worker.

At the present status of psychology these tendencies or attitudes or traits or aspects of personality as distinguished from capacity or ability, with a few exceptions, cannot be tested. The best that can be done is to obtain the judgment of persons familiar with the man in question. This judgment may be secured in most effective form in a rating scale. Such scales are utilized in various ways. Brief ones are used for estimating an applicant during an employment interview. Estimates of a systematic sort are obtained from previous employers, school teachers, or others who have been in touch with the applicant. Promotion from one department or job to another within an organization is a logical part of the personnel program and it is for this purpose, perhaps, that rating scales are at present most widely used. Many concerns have their employees rated periodically by their superiors. The results, on the one hand, indicate cases of malad-

justment where transfer or special training is requisite, and, on the other hand, serve to locate promotional material.

More Uniform Method of Expressing Opinion. Men are constantly observing one another and, from external behavior, inferring something regarding mental traits. These estimates are often built up almost unconsciously, but their effect is cumulative; when a person is asked for an opinion regarding another he may realize that he actually has such an opinion already formed. These opinions, however, are of somewhat dubious value in personnel problems, especially in the form in which they are most frequently available. If a man is asked what he thinks of a given executive or applicant, his answer usually involves some glittering generalities to the effect that he is a "good man" or he "does not take hold." These terms unfortunately are quite relative and mean radically different things to different individuals. Being a good man in the estimation of one person may be equivalent to mediocrity in the estimation of another. General impressions of this sort are likewise apt to reflect prejudice. If the rater has had some unfortunate experience with the person in question—for example, if he has encountered a single instance of carelessness—he is likely to impute the bad impression of this incident to the individual's entire personality.

Hence it is desirable to abstract from these prejudices and general impressions and obtain the estimates in more scientific fashion. This can be accomplished to a certain extent by rating the traits separately and then combining them into a final rating. If, for instance, one is considering tact, initiative, and leadership separately, his judgment will probably to a lesser extent reflect his general impression or the influence of a single dramatic incident than if he is giving a single figure which is to evaluate the individual as a whole. Separate consideration of the traits in this way obviates snap judgment and insures that all the raters will record their impressions in more systematic and, above all, in more uniform manner. The judges are somewhat less apt to disagree when rating traits separately than when estimating the individual as a whole, and if there is disagreement it is possible to analyze it because of the more uniform character of the whole technique.

Educational Value. Another aim of the rating scale procedure in organizations where periodic ratings are made is to educate

both the rater and the rated. The former comes to observe the latter more closely if he is required to rate him occasionally, and he becomes personnel-conscious. In addition to arousing personal interest in the man it leads the rater to observe him with reference to different traits and consider them separately. The natural tendency is to devote attention primarily to the man as a whole or to some outstanding aspect. It is easy to dislike a man's face and overlook his other good qualities. The rating scale calls attention to these other qualities and teaches one to observe them too. One may discover that, after all, the man is rather skillful, ingenious, and cooperative. On the other hand, the scale may call attention to the man's laziness which had been previously overshadowed by his affability. In this way one's final opinion of the man and one's whole attitude toward him may be very appreciably changed.

The use of rating scales in an organization likewise has educative value for the employee who is rated. He realizes that he is being judged in essential traits. This may encourage a certain amount of self-analysis and evaluation and he may seek to determine his weak points with a view to improvement. Sometimes the personnel department discusses these weak points with the employee. He may also realize that the ratings have something to do with his status in the concern, and hence they serve to motivate him to do as effective work as he can.

Check on Employees' Progress. Another purpose of the rating scale in some organizations is to give a periodic check on the employees' progress. Those whose development seems to be rapid and who are especially superior in certain traits may be considered as a source of supply for other higher positions within the concern and may be promoted. Others who seem weak in certain respects may be transferred to other departments for which they are better adapted. These readjustments may become desirable because it is often impossible to apply the rating scale effectively at the time of original employment. If personality traits could be measured by objective tests along with capacities, persons could at the outset be placed in that line of work for which they are best fitted. As this is not the case and it is often necessary to wait until superiors are acquainted with the employees before estimates as to personality traits are available, these later adjustments are frequently desirable.

Data to Meet Emergencies. If an organization has ratings of its employees on file, they may be useful in an emergency. Vacancies may occur unexpectedly and it may be desired to promote or transfer someone quickly. The usual practice in such a case is to consult other members of the firm as to their general impression of certain possible candidates for the position. If systematic rating scales are used at this time, they can scarcely be properly evaluated. As will be shown later in the chapter, it is necessary first of all to train the raters, have them rate a considerable number of employees of a given sort, compare the reliability of different raters, and if possible determine the validity of the ratings. It is often necessary to make corrections and run down special cases of discrepancy in which the rater apparently did not follow instructions. Only then can ratings yield their greatest value. If quick action is necessary in an emergency, this careful procedure is not feasible. Furthermore, the foreman's judgment at that time will deal exclusively with the man as related to the vacant job, whereas what may be wanted is a broader estimate of the man as the foreman "ordinarily thinks of him." Consequently, it is more satisfactory if individuals who might be possible sources of supply for other positions or departments are rated systematically in advance and the records filed for any contingency that may arise.

SELECTION OF TRAITS TO BE RATED

The mental characteristics to be included in a rating scale depend, of course, on the situation in which it is to be used. There is no one rating scale that is universally applicable any more than there is a universal test that can be used in selecting employees for every job. The mental make-up of a successful salesman is considerably different from that of a successful executive. Consequently, if a rating system is to be devised for salesmen, it should emphasize a different group of traits from those included in a similar system for executives.

Traits That Are Present in Varying Degrees. In selecting traits to be included in a rating scale, certain ones will be of dubious value. Such traits cannot be rated on a scale because they are not present in varying amounts. Such a trait, for instance, as loyalty is difficult to conceive in terms of more or less, for a

person is loyal or he is not loyal. The same thing may be said regarding honesty and various other traits in which it would be difficult to grade the individual on a scale. In such cases it is probably unwise to include the trait in the rating scale, because an effort to estimate it in varying amounts will only be confusing. If such traits are significant, the rater can be required merely to check one of two alternatives according to whether the person is loyal or disloyal, honest or dishonest.

Questionnaire. In determining what traits to include in a particular rating scale the logical procedure is to consult persons who are familiar with the occupation in question. Members of the staff who have been concerned with employing or promoting certain kinds of employees have doubtless been using some personal unsystematic consideration of such character traits as appear in rating scales. It is possible then to circulate to such people a questionnaire asking them to indicate the traits which they consider important for this particular type of work. Certain traits on which they agree fairly well may be considered of fundamental importance. Traits which are not so generally mentioned may be either discarded or made the subject of a conference which will bring out the reason why some members of the staff listed them while others did not. For instance, in devising a rating scale for salesmen each manager who was ultimately to use the scale was requested to submit independently a list of the traits which he ordinarily considered when estimating the value of a salesman. The most frequently mentioned traits were included in the final scale; they were as follows: experience, dominance, stamina, appearance and manner, enthusiasm, fluency, egotism, expansiveness [11, 189].

Interview or Conference. Possibly a better procedure than the foregoing is to interview the members of the staff who are to suggest the important traits. In working with mere lists, unless elaborate definitions are used, there are apt to be ambiguities in terminology. If a man writes on his questionnaire that a job needs "cooperativeness," it is impossible to tell whether he means merely willingness to do what one is told or whether he considers further the tendency to anticipate the needs of others and to govern oneself accordingly in advance. The real meaning which he attaches to the term can be brought out in a personal

interview. Sometimes this information is obtained in the interview conducted for the broader purpose of job analysis. (Cf. Chapter XV.) Sometimes a conference is desirable to iron out any apparent disagreements between the different members. It is well, however, to have each person commit himself independently in the first place, because in a conference those who speak first may exercise a certain amount of suggestion upon the others. If each man's unbiased opinion is a matter of record at the outset, the conference is valuable in determining the reasons for various disagreements.

This procedure was followed in developing a scale for rating interviewers in state employment services [23, 25]. Appointments were made with 25 members of the administrative staff, the importance of the rating form was explained to them, and they were asked to suggest qualities which they considered necessary for satisfactory performance as an interviewer. The result was a list of 70 traits or qualities and a tabulation of the frequencies with which each was mentioned. The list was reduced to 18 traits and after these were defined the experts were asked to indicate their relative importance by giving each one a rating from 0 to 10. By pooling these estimates a final list was selected.

In the case of salespersons it was possible to interview customers rather than sales managers [4]. The customers were questioned as to what traits they thought important, what irritated them, and how *they* would train a salesperson. In this way five traits were selected for the rating scale: interest in the customer, merchandising information, display of merchandise, courtesy, and alertness. This procedure is logical enough because after all the customers are the ones who will be influenced most directly by the traits of the salesperson.

Preliminary List from Which to Select. It has sometimes facilitated the procedure of questionnaire or interview to provide in advance a fairly exhaustive list of traits from which the persons consulted may select those which they deem important. People who find it difficult to recall, when requested, the traits which they consider in evaluating their subordinates may find such a list helpful. A typical list classifies a large number of traits under the following captions [10]:

1. General intellectual: ability, alert, bright, intelligent, keen, thinker (good).
2. Special intellectual: breadth, scholarship, initiative, originality, good judgment, resourceful, mature mentally.
3. Efficiency of performance: accurate, capable, efficient, responsible, thorough, careful, expresses self well.
4. Efficiency in attitude: ambitious, diligent, determined, energetic, enthusiastic, persistent, prompt, industrious, painstaking, will succeed, willing to work.
5. Social—indicating control of others: executive ability, forceful, influential, leadership, inspires confidence.
6. Social—moral: character (strong), altruistic, conscientious, dependable, earnest, faithful, honest, loyal, reliable, steady, sincere, unselfish, high ideals, trustworthy.
7. Social—attitude toward others: adaptable, agreeable, charming, cooperative, friendly, genial, kindly, modest, independent, popular, social mixer, tactful, winning personality, poise.
8. Miscellaneous: appearance, habits, etc.

WEIGHTING THE TRAITS

When the traits or qualities that are to be included in the rating scale have been determined, it is essential to consider their relative importance with a view to weighting them. It is possible that for salesmen tact may be twice as important as leadership. Just as when a number of tests are used for determining vocational aptitude the predictive value is raised if the tests are properly weighted, so the value of a rating scale is increased if the proper significance is attached to each trait. While in some cases the ratings on different traits could be compared with a criterion, this is seldom done. Ratings are less objective, quantitative, and reliable than tests so that this procedure would scarcely be worth the effort. Moreover, the criterion itself is often a rating. The relative importance of the traits is generally determined in rather arbitrary fashion by using the best judgment of those who are familiar with the occupation in question.

Frequency of Mention in Questionnaire. If a questionnaire has been circulated or job analysis interviews have been conducted, it is possible to note how many times each item is mentioned in the questionnaire or interviews and to weight it accordingly. If, for instance, 50 people consider leadership important and only

25 consider originality worth mentioning, leadership might receive a weight of 2, and originality a weight of 1.

Pooled Judgment. Another possibility is to submit the final list of traits to a considerable number of judges and ask them to distribute, say, 100 points among these traits, i.e., to assign a particular weight to each one according to its importance so that the total of the weights will equal 100. The weights assigned a trait may then be averaged. In cases such as the one mentioned previously, where the original traits were given a rating as to importance on a 10-point scale, it is feasible to average these assigned values. The weights finally adopted usually are not the exact averages but are rounded off to a convenient number frequently ending in 5 or 10. It is doubtful if a finer gradation than this is worth while because of the coarseness of the original judgments. In the rating scale for interviewers cited above, the average values assigned the traits differed so little that it seemed advisable to weight them equally.

According to Reliability. One other procedure for determining weights has occasionally been used. This involves weighting the items roughly according to their reliability. The methods just described involve rather the validity of the traits, i.e., their relative merits in predicting a criterion. The present suggestion assumes that, since there is considerable difficulty in ascertaining the validity, it is better to look for the most reliable traits. If the judges agree with one another fairly well on some traits and not on others, the former should receive more weight, not because they are more closely related to occupational proficiency, but because the ratings themselves come nearer to being a true index of the particular trait under consideration. It will be shown later in the chapter that traits which tend to yield some objective product by which they can be judged, such as salary or bank account, are estimated with greater reliability. Consequently, in lieu of actual measurements of reliability, more weight can be attached to such traits and less weight to those that are more subjective in character and hence presumably lower in reliability.

By way of illustration the weights assigned to the traits in a few rating scales will be cited. In the rating scale for salesmen mentioned above the following weights were adopted as a result of conference:

Experience	2	Enthusiasm	2
Dominance	3	Fluency	2
Stamina	2	Egotism	1
Appearance and manner	2	Expansiveness	1

A rating scale for Army officers was devised in 1918. After a considerable amount of study and revision the final list of traits and their weights were as follows:

Physical qualities	15
Intelligence	15
Leadership	15
Personal qualities	15
General value to the service	40

A rating scale for clerical workers in a large office force was devised. The items were weighted by consultation with ten division heads with long experience. The items were weighted in two ways—one for clerical duties involving only individual work and the other for clerical duties where supervisory work was entailed. The qualities with their weights for individual and supervisory work follow:

	Individual	Supervisory
Appearance	10	10
Ability to learn	20	20
Accuracy	25	10
Dependability	10	10
Speed	20	5
Cooperativeness	7.5	10
Constructive thinking	7.5	10
Ability to direct work of others	0	25

Incorporating Weightings in the Rating Blank. When the rating procedure is put into practical use, it may be arranged so that the rater gives no concern whatever to the weighting which is subsequently done by whoever evaluates the data, or the weighting may actually be embodied in the rating blank. In the case of salesmen just mentioned, the rater estimated each trait in the same terms and on the same basis. When the blank was scored, the rating in experience was multiplied by 3, that for stamina was multiplied by 2, etc. Similarly with the scale for clerical workers, each trait involved a graphic scale (*infra*) that

had the same maximum. If, then, the person was being considered for supervisory work, the different estimates were multiplied by one set of constants before being totaled, while if he was being considered for individual work they were multiplied by the other set of constants. In the Army rating scale, on the other hand, the rater used a master scale and considered his subordinates by comparison with other officers on the master scale. It was so arranged that in physical qualities he could assign a maximum value of 15, while in general value to the service he could assign a figure up to 40. In this way the weighting was done in the actual process of rating.

DEFINING THE TRAITS

Avoid Individual Interpretations. It is usually insufficient merely to present the name of a given trait and require a person to estimate somebody with reference thereto. If the scale mentioned "executive ability," the rater might construe it either from the standpoint of planning or from the standpoint of ability to get things done. If a scale involved the term "originality," this might be interpreted either as ability to work without supervision or as actual inventive capacity. Hence it is necessary to define the traits in more detail. It may even be advisable to make tentative definitions and revise them after preliminary use. In the Army rating scale, for instance, the item "intelligence" in the early form of the scale was described as "ease of learning, capacity to apply knowledge, ability to grasp and solve new problems." Later this definition became "accuracy, ease in learning, ability to grasp quickly the point of view of the commanding officer, to issue clear and intelligent orders, to estimate a new situation, and to arrive at a sensible decision in a crisis."

Typical definitions appear in some of the rating scales illustrated below. They are necessary in order that all the persons using the rating scale will have in mind exactly the sort of thing that is desired. Some workers in this field have suggested that in the final form of a scale it may be better to omit altogether the actual name of the trait and to include simply the definition. The idea is that if the trait is actually named, the rater may merely read the name and devote little attention to the detailed

definition, thus putting his own interpretation on the name. Omitting the name compels him to read the definition.

Objective Preferable to Subjective. In defining traits it is further desirable to do so as far as possible in objective rather than in subjective terms. Objective traits which represent reactions to impersonal things or situations or tasks, and which tend to yield some objective observable product, are rated more reliably than the opposite type. Some traits may be defined in either objective or subjective forms, and in such a case the former is to be preferred because the raters, taking a more objective attitude, will tend to be more reliable.

Consider leadership, for instance. An objective definition might be as follows: "Rate this executive in terms of the success he has shown in developing a loyal and effective organization by administering justice, inspiring confidence, and winning the cooperation of his subordinates." Here attention is called to actual objective accomplishment such as the organization he has developed, and to the way his subordinates react as a result of his leadership. A subjective definition of the same trait might read thus: "Rate this executive's initiative, force, self-reliance, decisiveness, tact, ability to inspire men and to command their obedience, loyalty, and cooperation." This definition calls attention merely to the subjective traits rather than to anything that results from their presence or absence. Again, personal appearance may be defined objectively as: "Consider how favorably he impresses people by his physique, bearing, and manner"; and subjectively as: "Personal attractiveness, cleanliness, neatness, and dress." Or a scale for salespeople might include an objective formulation such as: "Does this salesperson adapt her selling behavior to suit the particular type of customer or does she sell to all customers in the same manner?" as contrasted with a subjective: "Adaptability and flexibility." The presumption is that traits defined in the more objective manner will be more reliably rated.

With Reference to the Particular Situation. It is further desirable to define the traits with reference to the situation in which they are to be used. The definition that would be most satisfactory for rating an executive might differ somewhat from the definition that would be best for rating a subordinate. For

instance, in a scale for executives, foremen, and supervisors, cooperativeness is defined as: "Success in winning the cooperation of his subordinates, in welding them into a loyal and effective working unit." In a scale for other workers in subordinate positions this same trait is defined as: "His attitude of helpfulness toward others, his inclination to cooperate in manner as well as in act with associates and superiors." Again in the first scale initiative is defined as: "Success in doing things in new and better ways and in adopting improved methods in his own work." In the scale for workers it is defined as: "Success in going ahead with a task without being told every detail; ability to make practical suggestions for doing things in new and better ways." Thus it is necessary to consider definitions of the traits, as well as the actual traits themselves, with reference to the situation in which they are to be used, and to call attention to the particular things involved in the actual situation.

Where traits of rather general character are used, it is sometimes desirable, instead of defining them for the particular situation, to define them for a number of typical situations. Consider, for instance, a trait like "self-assurance." It is possible to name a number of situations in which it might have opportunity to manifest itself and to state for each a response that would apparently indicate a positive manifestation of this trait and another that would indicate a negative manifestation [16, 157]. The following situations are each followed by a possible positive and a possible negative response: (1) A new situation demanding response: positive—undertaking with readiness, carried out beyond demands; negative—excessive inquiry and waiting for directions. (2) Many tasks inviting response: positive—acceptance of many; negative—carrying a light load. (3) A task demanding preparation: positive—tendency to undertake without thorough preparation; negative—careful preparation. (4) Opinion asked: positive—readily given; negative—modestly withheld or qualified. (5) Contradicted when asserting one's own memory of an event: positive—denial of error; negative—acceding. The rater may be required to estimate the person in each of these hypothetical situations or to consider them in making a single estimate for the trait in general. At any rate, calling attention to these situations clarifies the matter.

MAN-TO-MAN RATING SCALE

Construction of the Master Scale. Four types of rating scales have been quite extensively used. The first of these involves man-to-man comparison and was patterned originally after a scale for rating salesmen developed by Scott. Its most extensive use was in a scale devised for rating Army officers in 1918. The unique feature is the construction at the outset of a master scale for each trait including a number of individuals of the same type as those on whom the scale is ultimately to be used. These individuals are selected at the average and extremes of the trait in question and their names written opposite the appropriate rating values. When a new man is to be rated he is compared with the men on the master scale and given a rating similar to the number assigned to the man on the scale whom he most resembles.

A typical master scale for rating minor executives is given below, with hypothetical proper names written in. Suppose, for example, that an executive is to use a scale for the six qualities indicated in rating the executives in the next grade below him. He is instructed to make out a list of some 15 or 20 such persons with whom he is well acquainted and to include in it some who are very good as well as some who are very poor in the characteristics in the scale. Next considering "appearance and manner" and disregarding every other characteristic, he selects the executive who surpasses all others in this respect. Suppose this is Smith; he writes this name on the first line of the master scale after the word "highest," as is done in the example below. Then he selects the one who is at the other extreme in appearance and manner and places his name, Briggs, on the line marked "lowest." He selects a third executive about midway between the two and enters his name on the line marked "middle" (Brown). He chooses another midway between Smith and Brown and one midway between Brown and Briggs. Weights for these five degrees of the trait had been determined previously—the highest degree 10, the next 8, etc. These five names from Smith to Briggs inclusive thus constitute the master scale for rating appearance and manner. Another master scale is constructed for leadership by writing in the names of five executives who cover

the range from highest to lowest, and similarly for the other four traits. When the blank is given to the executive who is to use it, no proper names are included; the rater supplies them himself.

The master scale used by one rater will seldom be the same as that used by another. It comprises persons who, in the opinion of the individual who will use the scale, cover the range of the trait indicated. It is possible for the name of the same person to appear on more than one master scale. The same man might, for example, be highest in the master scale for appearance and manner and in the middle position for ability to develop men. The only point is that the rater must think of the individual with reference to only one trait at a time. For most raters it is less confusing to have the names entirely different so that the same individual appears on only one scale.

RATING SCALE FOR MINOR EXECUTIVES

- I. *Appearance and manner.* Ability to inspire confidence and respect through his appearance and manner.

Highest, Smith	10
High, Jones	8
Middle, Brown	6
Low, Doe	4
Lowest, Briggs	2

- II. *Leadership.* Ability to elicit the cooperation of his colleagues and subordinates, to promote morale, and to develop a loyal and efficient organization.

Highest, Daniels	20
High, Callahan	16
Middle, Haskell *	12
Low, Ordway	8
Lowest, Clark	4

- III. *Organizing ability.* Ability to plan work wisely, to discriminate the relative importance of its different parts, and to delegate its administration properly.

Highest, Clyburn	20
High, Murphy	16
Middle, Hershey	12
Low, Eckert	8
Lowest, McCreary	4

IV. *Initiative.* Ability to get things done.

Highest, Pritchard	15
High, Titcomb	12
Middle, Goodwin	9
Low, Campbell	6
Lowest, Safford	3

V. *Ability to develop men,* by teaching them about their work, arousing their interest in it, and stimulating their desire to progress.

Highest, Cambridge	15
High, Varney	12
Middle, Rundle	9
Low, Clines	6
Lowest, Parsons	3

VI. *General value to the concern.*

Highest, Stillman	20
High, Cassiday	16
Middle, Sterling	12
Low, Brooks	8
Lowest, Thomas	4

Use of the Master Scale. After a rater has filled out the master scale in this fashion he can then use it for rating subordinates. In the foregoing example, if the first man to be rated is Adams he is compared with the five men on the master scale for appearance. If he seems most similar to Jones, he would receive a rating of 8. If he is somewhat inferior to Brown but not as bad as Doe, he would be rated 5. He would then be considered with reference to leadership in comparison with the five men on the master scale for leadership—Daniels, Callahan, Haskell, Ordway and Clark, and so on for the other traits. In each he would be given a rating similar to that of the man on the master scale whom he most resembles, or an intermediate rating where this is appropriate. The total of these ratings would represent his final standing and this figure would be used for the specific purpose for which the scale was constructed. In the present instance a maximum rating of 100 points is possible.

There are several advantages in this procedure of man-to-man rating. In the first place, it gets away from letter grades or "per

cents," which to many persons are variously associated with school grades. If the raters were requested, for example, to assign each subordinate a percentage between 0 and 100, they would be quite apt to think in terms of what the passing grade was in their school career. Some probably were accustomed to a passing grade of 50 and others to a grade of 70, and this would have the effect of sliding the "passable" workmen appreciably up or down the scale so that the results of different raters would not be comparable. In the second place, the master scale is a relatively permanent measuring device. One would not use a cotton yardstick for accurate physical measurements because it might shrink overnight. One's notion of a "75 per cent man" or a "B-grade man" may shrink or stretch in similar fashion depending on such causal things as the time of day, the digestive condition of the rater, or some compliment or insult that he has recently received. The master scale, however, should not shrink. Comparing the mental qualities of a group of executives with those of Smith, Jones, Brown, Doe, and Briggs today and making similar comparisons next week should yield comparable results. For while a "grouch" might lower one's opinion of the group that was being rated, it would also lower his opinion of Smith, Jones, etc. The ratings would all be relative to Smith, Jones, etc., regardless of the rater's mood.

RATING BY DEFINED GROUPS

Linear Scale. Another method which is frequently used involves not direct man-to-man comparison, but rating persons relative to the other members of a definite group. This group is used as a standard. The scheme is quite similar to that discussed in Chapter VI in connection with estimates by superiors used as a criterion. In that case, however, we were concerned merely with a single estimate for each individual as to his ability in the job, whereas now it is a matter of evaluating separate traits. (Cf. [17, 24].) A typical blank for such a rating scale is given below.

The introductory statement is practically self-explanatory, but in actual use it is well to go over it with the people who are to do the rating and insure that they understand what is required. The

traits would, of course, be defined in detail either on the rating blank or on a separate sheet. Such ratings may be quantified by measuring, in millimeters or some other convenient unit, the actual distance of the check marks from the left edge of the left column. The larger number will indicate a higher rating.

RATING SCALE FOR EXECUTIVES

Imagine all the executives of your acquaintance divided into five equal classes on the basis of their possession of each of the following traits, a highest fifth, a next highest fifth, a middle fifth, a next lowest fifth, and a lowest fifth. Now take the first man whom you are going to rate and, considering only his energy, compare him with these other executives. If you think, for instance, that he falls in the middle fifth, place a check on the line after "energy" in the column headed "middle fifth." If you think, on the other hand, that he is among the best 20 per cent in energy, check in the column at the extreme right. Furthermore, if, after you have located him in the proper column, you consider that he stands relatively high or low in that particular fifth, indicate accordingly by placing your check to the right or to the left. In other words, the farther to the right the mark is placed, the higher the degree to which the individual possesses the trait in question. Now take the same man and, considering him solely from the standpoint of initiative, compare him with the total group in that respect. Indicate your judgment by checking on the line for initiative in the same way. Proceed in the same fashion with the other traits.

	Lowest Fifth	Next Lowest Fifth	Middle Fifth	Next Highest Fifth	Highest Fifth
Energy.....
Initiative.....
Leadership.....
Tact.....
Organizing ability.....

In the form just discussed a different sheet is provided for each person who is to be rated. One individual at a time is con-

sidered. The procedure may be varied by arranging it so there is one sheet for each trait, as follows:

ENERGY

	Lowest Fifth	Next Lowest Fifth	Middle Fifth	Next Highest Fifth	Highest Fifth
Adams.....					
Andrews.....					
Briggs.....					

In this case the rater considers one trait at a time and goes through all the men with reference to this trait before considering the other traits at all. This latter procedure is theoretically preferable to the former. There is always a danger of considering general impression rather than the specific trait in question. (Cf. discussion of the halo effect, pp. 39, 388.) In the former method when the same man is evaluated with reference to the various traits in immediate succession, there is a danger that opinion regarding initiative will be influenced by the rating for energy made a moment before. In the latter method there is less danger of associating one trait with another, and the rater does not have to make such a special effort to abstract from other traits when considering a particular one. This method involves a little more preliminary clerical work in typing the names of persons who are to be rated. Some executives, moreover, may dislike the procedure of considering one trait at a time for all the men, because it is less natural and perhaps more difficult. However, the satisfactory administration of rating scales involves, as will be brought out later, some training of the raters and it is possible in this way to revise their habits as the method necessitates.

The division into five classes as in the above instance is not particularly essential. The following is another division that has been used. The part of the introductory statement which is similar to that in the preceding case is omitted.

Check in one of the columns running from "very high" to "very

low" to indicate the person's standing in the trait. Try to let the percentages guide you as to the number of check marks to place in each column.

LEADERSHIP

	4 Per Cent Very Bad	11 Per Cent Bad	21 Per Cent Poor	28 Per Cent Average	21 Per Cent Good	11 Per Cent Very Good	4 Per Cent Excellent
Adams.....							
Andrews.....							
Briggs.....							

Assignment to Classes. Some concerns have found rating scales of the foregoing sort too cumbersome and difficult for practical use. Instead of checking in the five columns and making fine gradations they merely have the rater assign a number from 1 to 5 to each man in each trait [12, 13]. These numbers may be defined somewhat as follows:

A central rating of 3 means that the employee meets reasonably satisfactorily the recognized departmental standards in respect to this trait. A 2 rating means that the employee is deficient enough in the trait under consideration so that he has had to be warned, criticized, or otherwise spoken to about it. A rating of 1 means that the employee is so seriously deficient in the trait that, if it is an important one, he is under consideration for transfer or dismissal. A rating of 4 means that the employee stands out above the general run of employees of the department in respect to this trait, while a rating of 5 means that the employee stands out so conspicuously from even the 4 men that he ought to be distinguished from them.

A large bank uses a method like this for a periodic rating of its employees, giving for each trait or characteristic a statement or set of questions calling attention to the main points and following this with the numbers 1 to 5. The rater rings one of these numbers. The following item is typical:

Consider how he applies himself to his work. Does he make his

daily tasks his main concern? Does he give his best and continuous effort to his work? Is he earnest, persistent, or easily distracted? Does he stick with his work till it is cleaned up? Does he use his time and ability to good advantage? Or does he tend to do as little as he can to "get by"? Does he need constant, occasional, or no supervision in order to get his work done on time?

1 2 3 4 5

The significance of the numbers is as above described. A similar procedure is followed for other items such as regularity of attendance, special knowledge or skill, tact, cooperation, ability to learn, responsibility, and general suitability. One scale is used for rating managers, another for the higher-grade clerical workers, and a third for the machine operators. The items in the different scales of course overlap to quite an extent. With such a technique the actual rating and also the subsequent recording of the results are more expeditious, but fine gradations of the estimates are impossible.

GRAPHIC RATING SCALE

Superiority to Other Methods. The two rating methods just described have certain shortcomings. The man-to-man scale proves rather cumbersome. It takes considerable time and effort to make up the original master scales satisfactorily, and even then the actual process of comparing individuals with the men in the master scale is tedious. Unless the raters are thoroughly "sold" on the value of the project, they are not inclined to devote sufficient time and effort to it.

The method of defined groups is less cumbersome, but in the linear form is almost too abstract for the average person who uses rating scales. It is a trifle difficult for the untrained to think in this fashion of the total range of a trait and to differentiate between the total range for "initiative," for instance, compared with the total range for "energy." Moreover, it is difficult to keep in mind the five or seven degrees of possession of a trait so that everyone will be judged on the same basis. If individuals are rated on different occasions they may unintentionally be rated according to a somewhat different standard. Merely as-

signing each person a number indicating in which of the five classes he falls is simple enough, but finer gradations are frequently desired.

The graphic rating scale has been devised to obviate some of these difficulties. It is less cumbersome and more expeditious than the man-to-man scale, because there is no master scale to construct. There is no necessity for carrying in mind standards as to total range or different degrees of a trait because these are all indicated by descriptive adjectives or phrases. The rater, moreover, can make as fine judgments as he wishes.

General Nature of the Scale. A graphic rating scale involves the name or definition of the trait, or a question embodying the trait, followed by a straight line a few inches long representing the distribution of the trait from maximum to minimum. Instead of the rater marking in arbitrary columns on this line, as in the method of defined groups, descriptive adjectives or phrases are placed along the line for his guidance. The adjectives range from those indicating a high degree of possession of the trait to those indicating a low degree. The rater checks at some point along this line as in the former method, but is guided by these descriptive adjectives. For instance, in rating a person as to social attitude, the line might have the following descriptive adjectives:

Constrained
and formal

Slightly
reserved

Meets one
halfway

Cordial and
informal

Extremely breezy
and informal

Construction of the Graphic Scale. The earlier discussion of the selection, definition, and weighting of traits is applicable to the graphic scale technique. It is quite common, however, instead of presenting a mere name or definition, to ask a question such as: "Does he strike out for himself in locating prospects?" The selection of the descriptive adjectives or phrases requires further discussion. For one thing, care must be exercised regarding the extremes that are selected. Occasions arise in which one word has several opposites and it is necessary to determine which is to be used. The word "ambitious," for instance, might be opposed to "lazy" or to "indifferent." The phrase "good leader" might be contrasted, on the one hand, with "too frequent

friction in his department" or, on the other hand, with "has to be led." In the one instance, leadership is thought of especially from the standpoint of maintaining harmony and, in the other, from the standpoint of actually telling people what to do compared with being told what to do oneself.

When extremes have been selected in this fashion, the intermediate phrases must conform to the extremes. If, for example, leadership is construed from the standpoint of harmony, the intermediate adjectives should deal with that general sphere, such as "obtains good cooperation" or "men dislike to work with him." In selecting extreme terms one should, moreover, avoid those that are so far from the average that they will never ordinarily be used. In rating ordinary workers there would probably be no place for the term "inventive genius," even though some aspect of originality was being considered. The extreme phrases usually are printed flush with the end of the line, although occasionally they are set in a little with the avowed purpose of suggesting that no one is perfect.

Effort should be made to select words that are as concrete and specific as possible. Terms like "very," "good," or "highly" should be avoided. It is much better to use something that connotes a definite situation. If an individual is being rated on sense of humor, it would convey to the rater a much more definite notion to say "often has to have jokes explained to him" than to say "poor sense of humor."

There is no fixed rule as to the number of descriptive adjectives or phrases that should be used. In general practice from three to five seem satisfactory. Three terms give opportunity for two extreme values and one intermediate or average value, while five terms facilitate slightly more detailed grading. Five terms are usually sufficient to give the rater an adequate idea of the distribution of the trait. If the type is not too large there is still ample white space between the phrases so that the rater can indicate intermediate ratings.

It is not always necessary that the adjectives be equally spaced along the line. In fact, there are cases in which they ought to be unequally spaced because some adjacent pairs may actually describe individuals who are more similar than those described

by other adjacent pairs. For instance, a set of four phrases for rating leadership might be distributed somewhat as follows:

Inspiring leader	Handles men well	Men have little confidence in him	Continued friction with subordinates
---------------------	---------------------	--------------------------------------	---

In this case the two intermediate phrases are more closely related to the respective end phrase than to each other. The terms "inspiring leader" and "handles men well" are both positive in character and somewhat related, while the other two are similarly related, both being somewhat negative. Hence the largest space is left at the middle of the line in accordance with the actual distribution of the trait in question. Ordinarily the neutral point of a trait which runs from good to bad should be located near the center of the line in conformity with a normal distribution.

As in the defined-groups scale, all the traits can be printed on one page and the worker be rated on each trait in succession, or the page may include only one trait on which all the workers are to be rated before the next trait is considered. The former system is more convenient, but the latter is preferable scientifically because it tends to minimize the halo effect. In the former, with a considerable number of traits on a given sheet, it is advisable to arrange them with the high extreme sometimes at the right and sometimes at the left. If this is not done and a person is rather superior in most traits, the rater will make his marks consistently along the right edge of the blank. Then if he comes to a trait in which the person is actually somewhat inferior, he is liable to continue the tendency to mark toward the right, or at least the impulse to continue will appreciably bias his judgment of the inferior trait. This halo tendency is the bugbear of rating technique, and the graphic method with all extremes at one end aids and abets this tendency, for the rater is inclined to make all his marks in one position on the blank. If the extreme values are staggered, it breaks up this tendency and makes him scrutinize each line a little more closely. A set of rules for the construction of graphic rating scales has been formulated by Guilford [8, 271].

Typical Graphic Scales. A few graphic rating scales will be given by way of illustration [1, 21].

GRAPHIC RATING SCALE FOR EXECUTIVES, DEPARTMENT HEADS, FOREMEN, AND SUPERVISORS¹

Consider his success in winning confidence and respect through his appearance and manner.

Inspiring	Favorable	Indifferent	Unfavorable	Repellent
-----------	-----------	-------------	-------------	-----------

Consider his success in doing things in new and better ways and in adapting improved methods to his own work.

Highly constructive	Resourceful	Fairly progressive	Routine worker
---------------------	-------------	--------------------	----------------

Consider his success in winning the cooperation of his subordinates in welding them into a loyal and effective working unit.

Capable and forceful leader	Handles workers well	Fails to command confidence	Frequent friction in his department
-----------------------------	----------------------	-----------------------------	-------------------------------------

Consider his success in organizing work of the department or unit, both by delegating authority wisely and by making certain that results are achieved.

Effective even under difficult circumstances	Effective under normal circumstances	Lacks planning ability	Inefficient
--	--------------------------------------	------------------------	-------------

Consider his success in making his department or unit a smooth running part of the whole organization; his knowledge and appreciation of the problems of the department.

Exceptionally cooperative	Cooperative	Not helpful	Difficult to handle	Obstructionist
---------------------------	-------------	-------------	---------------------	----------------

Consider his success in improving his subordinates by imparting information, creating interest, developing talent, and by arousing ambition.

Develops workers of high caliber	Develops workers satisfactorily	Neglects to develop workers	Discourages and misinforms workers
----------------------------------	---------------------------------	-----------------------------	------------------------------------

Consider his success in applying specialized knowledge in his particular field, whether by his own knowledge of ways and means or through his use of sources of information.

Expert	Competent	Uninformed	Neglects and misinterprets facts
--------	-----------	------------	----------------------------------

¹ After Scott and Clothier.

GRAPHIC RATING SCALE FOR INVESTIGATORS, SECRETARIES, SPECIAL WORKERS, AND OTHERS NOT CHARGED WITH SUPERVISION²

Consider the ease with which this employee is able to learn new methods; the ease with which he follows directions.

Very superior	Learns with ease	Ordinary	Slow to learn	Dull
---------------	------------------	----------	---------------	------

Consider the amount of work he accomplishes; the promptness with which he completes it.

Unusually high output	Satisfactory output	Only average	Limited output	Unsatisfactory output
-----------------------	---------------------	--------------	----------------	-----------------------

Consider the neatness and accuracy of his work and his ability constantly to maintain high workmanship in these respects.

Highest quality	Good quality	Mediocre	Careless	Makes many errors
-----------------	--------------	----------	----------	-------------------

Consider his energy and his application to the duties of his job day in and day out.

Very energetic	Industrious	Spasmodic or indifferent	Needs constant urging	Lazy
----------------	-------------	--------------------------	-----------------------	------

Consider his success in going ahead with a task without being told every detail; his ability to make practical suggestions for doing things in new and better ways.

Very original	Resourceful	Occasionally suggests	Routine worker	Needs constant supervision
---------------	-------------	-----------------------	----------------	----------------------------

Consider his attitude of helpfulness to others; his inclination to cooperate in manner as well as in act with associates and superiors.

Highly cooperative	Cooperative	Not helpful	Difficult to handle	Obstructionist
--------------------	-------------	-------------	---------------------	----------------

Consider his present knowledge of his work and of other work related to it.

Complete	Well informed	Moderate	Meager	Lacking
----------	---------------	----------	--------	---------

² After Scott and Clothier.

GRAPHIC RATING SCALE FOR CLERICAL WORKERS³

<i>Appearance.</i> Neatness of person and dress.	Appropriate	Neat	Ordinary	Passable	Slovenly
<i>Ability to learn.</i> Ease of learning new methods.	Very quick	Catches on easily		Needs repeated instruction	
<i>Accuracy.</i> Quality of work; freedom from errors.	No errors	Very careful	Few errors	Careless	Many errors
<i>Dependability.</i> How well can he be relied on to work without supervision?	Very reliable	Trustworthy	Usually reliable	Unreliable	
<i>Speed.</i> Amount of work accomplished.	Very fast	Rapid	Moderate	Slow	Very slow
<i>Cooperativeness.</i> Ability to work with others.	Cooperative	Falls in line	Difficult to handle	Obstructionist	
<i>Constructive thinking.</i> Ability to grasp a situation and draw correct conclusions.	Shows originality	Resourceful	Carries out suggestions	Needs detailed instruction	
<i>Ability to direct work of others.</i> Ability to direct and gain cooperation.	Gets maximum efficiency	Directs work without friction	Secures limited cooperation	Wastes man power	Antagonizes

The following are a few items from a graphic scale used for salesmen.⁴ Particularly to be noted is the effort to deal specifically with what the salesman does rather than with abstract qualities.

Does he strike out for himself in locating prospects?	Waits to be directed	Discovers some leads	Exceptional "nose" for prospects		
Does he impress people as being sincere?	All he says taken at face value	Usually inspires confidence	Gives impression of bulldozing	Arouses suspicion	
Does he put in full hours?	100 per cent attendance and punctuality	Commendable, better than the average	Satisfactory	Irregular	Very poor attendance
Does he use good judgment in handling complicated situations?	Acknowledged blunderer	Makes an occasional error	Can be depended on to use good sense	Exceptionally clever in handling situations	

³ After Bills.

⁴ From H. G. Kenagy and C. E. Yoakum, *The Selection and Training of Salesmen*, by permission of The McGraw-Hill Book Company, Inc., New York.

Does he dominate an interview?

Agrees with everything a prospect says	Easily thrown off the track	Usually guides conversation	Directs conversation; ready with a comeback	Completely dominates an interview
--	-----------------------------	-----------------------------	---	-----------------------------------

How carefully does he study each prospect, his needs and attitude?

Has poorly considered plans	Has loose plans for prospects	Knows all that is readily available	Makes careful plans for big prospects	Goes deeply into every prospect's affairs
-----------------------------	-------------------------------	-------------------------------------	---------------------------------------	---

To show the possibilities of such a scale in an entirely different field a few items from a rating scale for teachers are given [6].

Is he self-conscious or self-possessed?

Painfully self-conscious and ill at ease	Frequently embarrassed or flustered	Self-conscious at all times	Usually unmoved by actions or remarks with reference to himself	Always at ease; self-possessed
--	-------------------------------------	-----------------------------	---	--------------------------------

Is he alert or absent-minded?

Always wide awake and alive to present situation	Usually has his wits about him	Fairly alert	Frequently becomes abstracted	Head in the clouds; preoccupied
--	--------------------------------	--------------	-------------------------------	---------------------------------

Does he display a sense of humor?

Sees funny side of everything	Usually sees the funny side of things	Slow in response to the comic	Often has to have jokes explained to him	Takes everything literally
-------------------------------	---------------------------------------	-------------------------------	--	----------------------------

How popular is he with his students and associates?

Arouses repulsion; detested	Disliked	Arouses neutral attitude	Liked	Popular favorite
-----------------------------	----------	--------------------------	-------	------------------

Is he prejudiced or fair-minded?

Partial and prejudiced; intolerant	Opinionated; has well-developed dislikes	Tries to be fair; usually just	Always impartial and fair-minded
------------------------------------	--	--------------------------------	----------------------------------

Scales like the above have proved valuable in certain organizations. Just as with mental tests, however, there is no guarantee that they will work in the original form in all concerns. They should be scrutinized by members of the staff to determine whether they will probably meet the needs of the particular situation. It is possible that some of the traits indicated will be of little importance and that others should be added. But the foregoing scales are typical and the methods used in their development can be employed in any similar project.

Scoring the Blank. The actual score represented by each mark on the rating blank is its distance from the right or left edge of the blank. This can be measured directly with a ruler in mil-

limeters or some other small unit. A simpler procedure is to use a celluloid stencil ruled into 5 or 10 vertical columns with the width between the extremes equal to the length of the lines used in the rating scale. This stencil can be placed over the blank and the check marks read directly according to the column in which they appear. Some psychologists make the columns near the center wider so as to give a more nearly normal distribution of the ratings. It is even possible to weight the different traits while scoring them. If, for instance, one trait is to receive twice the weight of the other, the stencil for the former may comprise 10 columns and for the latter only 5. In this way if the columns are numbered from left to right a check mark near the extreme right will receive a rating of 10 in one case and 5 in the other. On the other hand, the same stencil may be used for all the traits; if they are to be weighted unequally the resulting numbers are multiplied by the appropriate weight.

In using a graphic rating scale some raters are inclined to check directly above the phrase and fail to take advantage of the intermediate space. In some projects this is encouraged by providing parentheses for the check marks as in the following items:

SUGGESTIVE SELLING

()	()	()	()
Skillful, effective use of suggestive selling	Uses suggestive selling well	Suggestions not made in an effective manner	Irritates customer by tactless use of suggestions

MANNER

()	()	()	()
Discourteous and indifferent at times	Condescending, insincere manner particularly toward some customers	Secures confidence of customers; arouses interest in merchandise	Manner a definite asset in selling; interested in pleasing customer

Arbitrary scores are assigned to the different brackets. With this procedure there is no great advantage to the linear arrangement of the items. In fact, sometimes a scale is made out in linear form with the items on the line not arranged in sequence, as in the following:

WILLINGNESS TO WORK

()	()	()	()	()
Manifests eagerness to work by consistent hard work	Sluggish and slow to serve many customers	Does as much as the average worker	Serves customers promptly	Only wants to get by. Stands around and lets others work

CHECK LIST

Checking Descriptive Phrases. The next step is a transition to what may be considered a fourth type of rating scale—some form of check list with no linear arrangement. After a trait is listed, the descriptive statements follow in a column, with space for checking each. The following is a scale of this sort, entitled "Ohio Youth Personnel Record," devised for rating young people on N.Y.A. projects.

GENERAL TRAITS⁵

WORK REGULARITY	Check Here
Regularly reports as scheduled	
Notifies when absence is necessary	
Takes necessary absences without notification	
Frequent absences with notification	
Frequent absences without notification	

PUNCTUALITY	Check Here
Always late	
Frequently late	
Occasionally late	
Always prompt	

ADAPTABILITY	Check Here
Works and adjusts well on a variety of jobs	
Dislikes changing jobs, but does well	
Works well only on jobs he likes	
Likes change, but does not work out well	
Dislikes change, does not do well	

DEPENDABILITY	Check Here
Works steadily whether supervised or not	
Requires only occasional encouragement	
Works in spurts	
Must be given frequent attention	
Requires constant supervision	

THOROUGHNESS	Check Here
Product requires completion by others	
Product is sloppy, inaccurate	
Quality of product is variable	

⁵ Courtesy of National Youth Administration of Ohio.

Product is acceptable
 Product is outstanding

SUPERVISORY COOPERATION Check Here

Takes criticisms as a personal insult
 Resents suggestions
 Listens to suggestions but fails to comply
 Follows suggestions willingly
 Asks for criticisms and suggestions

COOPERATION WITH OTHER YOUTH Check Here

Is belligerent, quarrelsome
 Is domineering and bossy
 Does not mix with others
 Is pleasant only when he can gain by it
 Is pleasant and considerate consistently

INITIATIVE Check Here

Does only what is required
 Just doesn't see new things to do
 Shows poor judgment in trying new things
 Sees new things to do but asks first
 Does new things properly without being told.....
 Anticipates things to be done

This type of check list is coming into quite general use. The halo effect is somewhat minimized by the column arrangement, but even here it is advisable to put the favorable alternatives sometimes at the top and sometimes at the bottom of the group.

Miscellaneous Check List. A further extension of the procedure abandons altogether a listing of traits and provides merely a series of unclassified statements to be checked. Weights or scale values for these statements are derived [19]. One such project for rating salesmen will be described. The list comprised statements similar to those which are often made informally in evaluating a man, such as: "He is always on the job" or "He needs to be pepped up occasionally." Techniques have been developed in connection with attitude measurement whereby statements like these may be "scaled." Success in selling may be considered as a continuum, and these statements may be thought of as scattered along this line. The statement, "He is always on the job," is obviously pretty well toward the favorable end of the continuum,

while "He is in a rut" would be pretty close to the lower end of the scale. If the former is checked by a rater the salesman obviously should receive a few more points than in the case of the latter. The problem is to determine scale values which may be attached to such statements. Totaling and averaging the scale values of the items checked gives a notion as to about where one stands on this continuum of selling ability.

The empirical problem is to compile a list of statements like the above, then to select the best ones and derive the scale values for them. We can give only a hint of the rather involved statistical procedure. In the present instance 1000 statements were secured from sales managers and others. They were sorted by 14 judges who were familiar with the job requirements into 7 piles on the basis of their indication of effectiveness on the job. The statements on which there was not fairly good agreement among the judges were discarded immediately, leaving 132 items. These were scaled tentatively by Thurstone's method (cf. [8, 217 ff.]). The statistical technique is beyond the scope of the present discussion. Six hundred and fifty salesmen were then rated by from 2 to 5 of their superiors, who checked each statement as plus, minus, or doubtful. Each item was next analyzed statistically. Items with the smallest variability were selected on the ground that such statements were rather specific. Items were selected where there was a large difference between the average score, based on all the items, of the men for whom an item was endorsed and those for whom it was not endorsed. These data as well as the percentage of the group endorsing each statement were plotted against the tentative scale values. Curves were fitted by inspection and where necessary the scale value was corrected accordingly. In this way about 50 items were selected for the final check list. A few of them with their scale values follow:

	Scale Value
1. He is somewhat in a rut on some of his brand talk.	32
2. He tends to keep comfortably ahead of his work schedule.	56
3. He is a good steady worker.	46
4. He is weak on planning.	29
5. He is making exceptional progress.	69

In actual practice the scale values of the items checked by a rater are totaled. In the present example the reliability of the scale was .90.

JUSTIFICATION OF THE RATINGS

A device that is sometimes used to increase the care with which the ratings are made and, it is hoped, their reliability is to leave a space after each item in which the rater may justify his judgment. It is prefaced by a statement such as: "Give specific illustrations of behavior on which your judgment is based." A widely quoted scale of this sort is one devised for the American Council on Education for rating students [3]. It urges the rater to describe significant performances in support of his judgment in order to show specifically how the students manifest the qualities mentioned. The following statement was given in support of favorable ratings on the item, "Does he need constant prodding or does he go ahead with his work without being told?": "In a course in Elizabethan drama he voluntarily built to scale a model of the Black Friar's theater and the Fortune theater based on the work of Tambers and demonstrated Elizabethan methods of staging several of the plays read."

When a rater is required to support his judgments in this way it is natural that he should take them more seriously. There is of course the danger that filling out this more detailed blank will make too much of a demand upon his time so that he will not do it carefully. Granted, however, adequate cooperation on the part of the rater, this procedure has a good deal to recommend it and in some instances it has been claimed to increase the reliability of scales appreciably.

RELIABILITY OF RATINGS

Conformity to Normal Distribution Curve. In dealing with ratings as in dealing with tests it is desirable as far as possible to determine their reliability and validity. Some notion as to their reliability may be obtained by making a distribution curve of the ratings assigned by a given person and noting whether the distribution is normal (cf. p. 192). The presumption is that traits of this sort are distributed in about the same fashion as are the various mental capacities, and hence that correct ratings of

these traits will yield a normal distribution. The expectation is, for instance, that executives who are fairly capable at developing subordinates will predominate, and that, as we go toward the extremes of those who discourage and misinform their subordinates and those who develop men of exceptionally high caliber, the numbers decrease. Hence, if the ratings made by a certain person differ considerably from the normal type of distribution, we may suspect that something is the matter. If the curve is skewed with a predominance of high or low ratings, it is probable that he is using too strict or too lenient a standard. If the curve is steep with very little scatter, he is probably not making sufficiently fine distinctions between the men and is not considering the whole range of the trait. If we suspect that the rater is too strict or too lenient, it may be possible to have him rate a group that is known to be mediocre and see if he assigns them the same extreme values. In cases such as the foregoing it is well to confer with the rater and show him his tendencies. After such a conference he may rerate the men and perhaps obtain something like a normal curve. Possibly this procedure will result in more reliable ratings.

Correcting Skewed Ratings. These facts also suggest the possibility of correcting the original ratings statistically. The procedure discussed in Chapter VI for making heterogeneous criteria comparable is applicable in this connection. In that case, it will be recalled, the estimates made by each foreman were converted into terms of the total distribution of ratings made by that foreman, i.e., into standard scores. In the present instance, where all the ratings are on incommensurable traits and considerable unreliability is to be expected, there may be some doubt as to the value of such refined statistical procedure. A scheme that has sometimes been used consists of taking a considerable number, perhaps 50, of ratings made by a given individual, arranging them in order from best to worst, and calling the best 10 per cent A, the next 20 per cent B, the middle 40 per cent C, the next 20 per cent D, and the lowest 10 per cent E. Subsequent ratings made by this individual may be converted into these same letters. Thus, after all, whether the rater takes a high or low standard, those whom he puts relatively high will receive an "A" rating, indicating desirable possession of the trait,

while those with relatively low ratings will receive a grade of "E."

Agreement of Raters with Each Other: Man-to-Man Scale. A more direct approach to the reliability of rating scales may be made by noting the agreement of a rater with other raters or with himself. The former will be discussed first. Such an evaluation was made of the reliability of the officer's rating scale mentioned previously [20]. When 300 men who had been in an officers' training school together from two to three months made up master scales and rated one another, there was marked disagreement in the standing of an officer in the opinion of his fellows. The results for ten typical men are given in Table 39. One column gives the lowest rating each man was assigned by

TABLE 39. VARIABILITY OF RATINGS MADE BY FELLOW OFFICERS WITH MAN-TO-MAN SCALE⁶

Officer	Lowest Rating by Fellow Officer	Highest Rating by Fellow Officer
A.....	52	80
B.....	38	67
C.....	66	92
D.....	36	73
E.....	53	87
F.....	48	83
G.....	43	77
H.....	43	71
I.....	39	75
J.....	32	65

his fellows and the other column the highest rating he was given, out of a maximum of 100 points. Officer A, for example, was rated as low as 52 points by one of his fellow officers and as high as 80 by another; B had ratings as low as 38 and as high as 67. This indicates a considerable chance that a man may be located at some distance from his true position.

In other groups, where the raters had considerable training and discussion and then rated all of their fellows whom they felt

⁶ After Rugg.

competent to rate, the results were somewhat similar. Most of the individuals varied as much as 30 points in the ratings they were given by the other members of the group. It was estimated that the chances were not over four to one that any rating would be within 14 points of the true rating.

The foregoing appears to be the only systematic investigation of the reliability of the man-to-man type of rating scale. It is unfortunate that other such studies have not been made, for it is hazardous to generalize on the basis of this single one. If this low degree of reliability proved to be general, we certainly should discount the use of man-to-man rating scales. As suggested earlier, they are rather cumbersome for industrial use unless adequate cooperation can be secured by the raters and adequate time be devoted to the process. It is not clear in the above study whether this was the case. On theoretical grounds the man-to-man scale has considerable merit because the ratings are not subject to daily fluctuations in the attitude of the rater.

Agreement of Raters with Each Other: Graphic Scale. A somewhat more encouraging result has been found when the reliability of the graphic rating scale has been studied in this manner [18]. Results are available in which the same workmen were

TABLE 40. CORRELATIONS BETWEEN RATINGS MADE BY PAIRS OF FOREMEN WITH GRAPHIC SCALE⁷

Foremen	Correlation
A and F.....	.33
H and D.....	.78
J and K.....	.82
L and M.....	.63
N and O.....	.80
N and P.....	.75
O and P.....	.84

rated by two different foremen. The agreement between the two raters was computed by the usual correlation method. The coefficients for different pairs of foremen appear in Table 40. With the exception of Foremen A and F, there is a fairly high agreement between the different pairs. These two men were shown in other studies of their ratings to be rather inconsistent;

⁷ After Paterson.

when rating men on different occasions they did not agree well with themselves.

In an investigation where instructors rated students with a graphic rating scale on seven different traits, pairs of instructors did not correlate with each other as highly as did the foremen in the preceding study. The average of such correlations was .41 for one group of students and .38 for another group [15].

Agreement of Rater with Himself. Another approach to the reliability of ratings may be made by considering the agreement of the rater with himself. This is analogous to determining the reliability of a test by giving it twice. An individual's ratings on different occasions can be compared, noting merely whether he assigns approximately the same average rating in each instance. This indicates whether he keeps about the same subjective standard. Or successive ratings may be correlated. Results for a few foremen appear in Table 41. The table gives the correlations between the first and second ratings made by a given foreman and also between his second and third ratings. The first

TABLE 41. CORRELATIONS BETWEEN SUCCESSIVE RATINGS BY THE SAME FOREMAN WITH GRAPHIC SCALE⁸

Foreman	First and Second Ratings	Second and Third Ratings
B.....	.91	.96
H.....	.88	.92
C.....	.85	.86
G.....	.84	.92
I.....	.84	.90
D.....	.82	.90
F.....	.62	.66
E.....	.60	.82
A.....	.52	.88
Average.....	.76	.87

few men in the list obviously are quite reliable even at the outset, whereas the last few are not so reliable. These latter, however, improve considerably so that there is a closer agree-

⁸ After Paterson.

ment between their second and third ratings. The averages for the two columns show the general tendency for greater reliability to characterize the later ratings. This doubtless reflects the practice which the rater has had and substantiates the need, to be brought out presently, for giving raters definite training and practice.

In the study of instructors' ratings previously described, whereas the correlations between the different pairs of instructors averaged only .41, the correlations between two ratings by an instructor averaged .60. This suggests that each instructor was somewhat consistent with himself, but that the various instructors were basing their judgments on different criteria. With a graphic scale for rating executives which required the rater to justify his judgments, the reliabilities of the same rater over a six-month interval ranged from .58 to .79 for managers rating the men under them, and from .58 to .71 for supervisors rating the managers.

The check-list type of scale for salesmen mentioned above was made up in two forms and when these were checked by the same individual a month apart the correlation between the two was .85. This reliability is quite high as rating scales go. It may reflect the careful statistical work which was done in selecting and scaling the items for the check list.

The above discussion indicates that rating scales are none too reliable. In general, the reliabilities reported are lower than those obtained with tests. For one thing, the rating procedure is less objective. Other factors conducive to poor reliability will be mentioned in connection with the discussion of errors (*infra*). But specific scales differ in reliability and thus it is advisable to investigate this aspect before using a scale in personnel work.

We note also that raters themselves vary considerably in the reliability of their estimates. We might expect such differences, just as some persons have higher intelligence or quicker reaction time. Rating doubtless involves some native aptitude and some acquired facility. We shall discuss the latter aspect in connection with training raters. If an organization is using rating scales in a continued program it may well locate the more reliable raters and attach more weight to their results. Certain

raters may prove to be especially reliable or unreliable in rating certain traits.

VALIDITY OF RATINGS

The validity of a rating, i.e., its correlation with a criterion, is as important as its reliability but is usually more difficult to determine. In many instances the criterion itself is a rating so that correlation is impossible. It is often difficult to set a production criterion in occupations such as executive work where rating scales are specially used. There have been, however, a few studies of the validity of rating scales which may be cited by way of illustration.

Army Rating Scale. One of the items in the officer's rating scale used in 1918 was intelligence. Many of the officers who were rated for this trait also took the Army intelligence test [20]. When individual intelligence scores and ratings were correlated in fifteen different groups the coefficients averaged less than .05. The officers making the ratings had had little experience and training. After they had received further instruction in the technique, the correlations averaged .15. The training apparently produced a slight improvement.

If, however, the ratings made by several officers on the same man were pooled to obtain an average rating for that man and these average ratings in intelligence were then correlated with measured intelligence, the coefficients for three different groups were .48, .51, and .36. The pooled judgments are manifestly more valid than the individual judgments. The conclusion was drawn that "the averaging of three or four judgments would locate a person in his proper fifth of the scale."

Pooled Judgments. The preceding point should be emphasized, viz., the greater validity of composite judgments. In physical experiments we approach more nearly the correct answer by combining a number of measurements so that the chance errors tend to counteract one another. Similarly in ratings, if several persons do the job independently their averages should dispose of some of the idiosyncrasies of the judges. In general, we should hesitate to take the ratings of a single judge as the final answer. It is desirable to have several make the ratings and then pool the results.

Ratings of Salesmen. With a rating scale for salesmen (p. 372) some of the items were validated by comparison with annual earnings. Of those who were rated as having an "exceptional nose for prospects," 16 were very good salesmen—i.e., they earned over \$5000; 10 were good, earning \$2000 to \$4000; 4 were mediocre (\$1000 to \$2000); and none were poor. On the other hand, for those rated at the other end of the scale as having to "wait to be directed," the numbers in these same four salary groups were respectively 0, 2, 7, and 8. With another item concerning how well the salesman studies his prospect, of those who were rated in the best 30 per cent on the scale, 91 per cent were classed in the successful group, whereas only 57 per cent of the entire group were so classed.

Another project with salespeople showed a fair degree of validity. The criterion consisted of ratings by the educational director, that is, an overall rating which itself is subject to some limitation. The actual detailed rating scale in graphic form included vocabulary, voice, adaptability to buyer, dexterity in demonstration, character, knowledge of line. The correlation of this scale with the criterion was .66 [7].

In another project of this sort the criterion was again an overall rating based on total sales experience. The detailed ratings were made by customers and by service shoppers, that is, people who went about incognito making purchases for the purpose of sizing up the salesperson. The items in the scale included interest in customer, merchandise information, display of merchandise, courtesy, and alertness. These items were correlated with the criterion and then weighted in a regression equation. The multiple correlation of the weighted sum of the items with the criterion was .85 [4].

A concern using rating scales should strive, where it is possible, to make some determination of their validity. Where production or salary or some fairly objective criterion is available, this can readily be done. In some cases more indirect criteria are available, such as membership in technical or other organizations, holding office therein, or being listed in *Who's Who*. It is also possible to follow up the individuals after a period and compare their later success with their earlier ratings. Unfortunately, in many instances it is necessary to be content, for a

time at least, with a study of the reliability of ratings with no consideration of their validity.

SOURCES OF ERROR IN RATING PROCEDURE

It is in point to consider possible sources of error in rating procedure with a view to improving reliability and validity. We have already mentioned the locus of considerable error in the man-to-man scale, namely, careless construction of the master scale. Inadequate training of raters was mentioned as another source. There are, however, other factors that may introduce errors into any of the rating procedures discussed above.

Comparative Reliability of the Estimates of Different Traits. One of the factors that must be considered is comparative reliability of the estimates of different traits. It has been found that some traits are more difficult to estimate than are others. The results of two studies bring out this point. In one study [9, 79], a group of individuals was rated by several judges with reference to a considerable number of traits. The variability of the judges or their disagreement with one another was computed for each trait. The results are shown in Table 42. Group I was rated by twelve judges and Group II by five judges. To make the two studies more comparable, the average disagreement of the judges on all traits was taken as 100, and the index of disagreement on each separate trait was reduced to the ratio of that index to the average disagreement. Figures smaller than 100 indicate closer agreement, and figures larger than this indicate greater disagreement, than average. The traits in the table are arranged roughly in order of closeness of agreement. If we consider the average column the traits may be grouped into the three classes indicated, showing close, fair, and poor agreement. The most noticeable thing is that the "close-agreement" traits are somewhat more objective in character than are the "poor-agreement" traits. By "objective" is meant that they tend to yield objective results or products such as inventions, books, positions, salary, bank account, property owned, and the like. A man's efficiency or originality or perseverance is apt to yield objective products to a greater extent than is his integrity, cooperativeness, or kindness. The latter traits manifest themselves more in a social situation, and after they have been manifested there is

TABLE 42. AGREEMENT OF JUDGES IN ESTIMATING VARIOUS TRAITS^a

Trait	Group I	Group II	Average	Classification
Efficiency.....	75	92	83	Close agree- ment. Average 88
Originality.....	95	77	86	
Perseverance.....	75	101	88	
Quickness.....	90	88	89	
Judgment.....	100	78	89	
Clearness.....	104	75	90	
Energy.....	75	109	91	
Will.....	85	98	91	Fair agree- ment. Average 100
Mental balance.....	110	81	96	
Breadth.....	100	92	96	
Leadership.....	90	103	96	
Intensity.....	85	113	99	
Reasonableness.....	115	86	100	
Independence.....	104	98	101	
Refinement.....	90	116	103	
Physical health.....	115	92	103	
Emotions.....	120	91	105	
Courage.....	100	119	109	Poor agree- ment. Average 117
Unselfishness.....	115	106	110	
Integrity.....	104	130	117	
Cooperativeness.....	125	113	119	
Cheerfulness.....	130	112	121	
Kindliness.....	120	125	123	

nothing objective to show for it. The objective traits more frequently involve reacting to things rather than to persons.

In another more exhaustive study of personality terms, especially those used in recommendations, somewhat similar results were found [10]. Eighty terms were classified according to the agreement between raters who used them. Somewhat the same trend is manifest. The alphabetical list of those on which there is greatest agreement in rating men begins with "ability, adaptable, breadth, dependable, diligent, expresses self well, hard-working, industrious," while the list on which there is least

^a From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, New York.

agreement begins with "alert, ambitious, cooperative, bright, character (strong), charming, cheerful, dignity." The traits in the first list are obviously more objective in the above-mentioned sense. When the ten most objective traits, as far as could be judged, were selected, the average index of disagreement (percentage of "maximum random disagreement") was .55, whereas with the less objective traits it was .70. This was with men rating men. With women rating women, on the other hand, the difference was negligible. This brings out the necessity, in this whole procedure of evaluating rating scales, of taking account of sex differences. This demonstrated greater reliability of objective traits substantiates the point made earlier that the traits should be defined in objective terms as far as possible. It further indicates the desirability of selecting for the scale traits that have this more objective character.

Desirable vs. Undesirable Traits. A study was made of the length of time required to make a rating of different traits [5]. It developed that a somewhat longer time was spent in estimating undesirable traits such as recklessness or obstinacy than in estimating desirable or neutral ones. The criterion of desirability consisted of previous estimates on these traits by the rater, who subsequently judged a sample of persons on these same traits. The data are not presented in such form that it is possible to determine the statistical significance of the trend, but the trend appears fairly consistent when analyzed from several angles such as judging friends vs. oneself or judgments above the average vs. below. The differences are of the order of 4 or 5 per cent. These results suggest that a rater might well vary his tempo when making his judgments. Perhaps he would do this naturally, but it might be worth while to call it to his attention as part of his training, and urge him to spend more time on his ratings of undesirable traits. The results, however, have no implications for the type of trait that runs the whole gamut from desirability to undesirability.

Halo Effect. Another source of error that is common in rating procedure has been called the *halo effect* [25]. This is the tendency to allow the general impression of the individual to color very markedly the evaluation of specific traits. If a man impresses us favorably either in a general way or by virtue of some

particular aspect of personality, or perhaps by some happy incident in our contact with him on the golf links, we are prone to invest his personality with a halo that sheds a luster upon his various traits and leads us to overestimate the desirable and to underestimate the undesirable in his personality. Conversely, if our general impression is unfavorable, this leads to underestimation of many of his desirable traits, and vice versa. The conventional halo is favorable in nature, but the halo in rating works both ways. If one is estimating a person's height he is little influenced by prejudice or by his general impression of that individual in other respects, but if it is a question of tact or industry or cooperativeness there is considerable danger of this error.

This same tendency was noted earlier (p. 39), where it was found that in estimating traits from photographs high correlations existed between such traits as humor, perseverance, kindness, courage, and intelligence. A person who looked as if he possessed a high degree of one of these looked as if he possessed a high degree of the others.

This halo effect is almost universal in rating procedure. It can be demonstrated easily in any laboratory class in which students use a rating scale in evaluating a group of acquaintances. It is simple to take the standing of those acquaintances in one trait and correlate it with their standing in another trait. If appreciable correlations are found, this may indicate a halo.

A few investigations of this general sort may be cited. A group of officers rated a large number of aviation cadets on the standard officer's rating scale. Correlations were computed between the different traits. The following correlations are interesting: intelligence and physical characteristics, .51; intelligence and leadership, .58; intelligence and personal qualities, .64. These are higher than one would expect the actual theoretical relation between the traits to yield. Experimental studies of intelligence in comparison with various measures of physical qualities, such as stature, strength, and agility, have shown the relation to be slight. It is evident that the officers in making their ratings fell into this common error of the halo.

A determination of the magnitude of the halo effect was made with the data from two teachers who had rated the same group

of pupils in seven traits [24]. For each pupil a composite rating of the seven traits was computed—one composite rating for each teacher. These composites indicated, as it were, the teacher's general impression of the pupil; and the more closely any given trait correlated with the composite, the greater was the effect upon that trait of the halo of general impression. The ratings on a given trait made by the two teachers were next correlated to determine, for instance, how well they agreed in estimating honesty. Then, by the technique of partial correlation (cf. Chapter IX), this same correlation was determined with the effect of the two composite ratings constant. The extent to which this partial correlation was lower than the original correlation showed how much the halo had raised the intrinsic relation between the ratings of the two teachers. These two sets of correlations are shown in Table 43. For example, the two

TABLE 43. MAGNITUDE OF HALO EFFECT IN CORRELATIONS BETWEEN RATINGS BY TWO TEACHERS¹⁰

Trait	Original Correlation	Correlation with General Impression Eliminated	Difference
Honesty.....	.47	.19	.28
Obedience.....	.39	-.04	.43
Courtesy.....	.41	.11	.30
Orderliness.....	.19	.10	.09
Cleanliness.....	.47	.55	-.08
Sportsmanship.....	.36	.00	.36
Promptness.....	.45	.09	.36
Average.....			.25

teachers apparently correlated to the extent of .47 in estimating honesty, but the intrinsic relation between their ratings, abstracting from general impression, was only .19, a difference of .28. Similarly, with all the other traits except cleanliness, the partial correlation is lower than the original. The average of these differences, .25, indicates roughly the magnitude of the halo effect in this particular situation.

¹⁰ After Symonds.

Avoiding Halo. The method adopted in many rating scales, of dealing with one trait at a time, is designed among other things to obviate this halo effect. It aids the rater in abstracting from the other traits while evaluating a given one. If he rates a man in all the traits in immediate succession, the effect of one is quite apt to influence another, and the general impression to influence them all. If he rates all the men on a single trait before considering the next trait, he tends to take an attitude of comparing the men with one another in one respect rather than considering the same man simultaneously in all respects. Even then, however, this halo effect is often present. The effort to define the traits under consideration more carefully and in objective terms will aid in directing the attention of the rater to the specific trait under consideration and away from general impression. Making him justify his rating insures more care and more specific consideration of the trait in question. In graphic scales the items may be staggered; that is, the favorable statements may be at the left end of the line for some items and at the right end for other items. In training raters particular stress must be laid on this halo error, for it is one of the most insidious difficulties in the rating scale technique.

It has been pointed out, however, that in one sense a certain amount of halo is legitimate [2]. This is exemplified by the general impression that goes with a position for which a person is being considered. If, for example, people are being rated with reference to an executive position it is legitimate to consider such characteristics as voice, appearance, poise, freedom from bias, and ability to plan and organize, against the background of the job that is to be filled. This does not mean that the rater confuses one trait with another, but that the estimate of each trait is slightly influenced by consideration of the job as a whole. This procedure is legitimate. Some projects even include an overall item like "general suitability for the job."

Length of Acquaintance. Another factor to be considered in ratings is the length of acquaintance. Obviously, if a supervisor has known a subordinate only a few days he can give only a rather poor evaluation of his various traits. Statistical support for this thesis is available. A study was mentioned previously in which instructors rated students on seven traits with a graphic

scale [15]. Three ratings of each student were taken at random and the average was correlated with the average of three other random ratings. In a sample of seniors, the reliability in this sense varied between .34 and .78, with an average of .67, whereas with the other students it ranged from .31 to .50, the average being .40. Assuming that the teachers were better acquainted with the seniors, the better-known students were rated with more reliability. In looking toward greater reliability of ratings adequate acquaintance should be insured.

On the other hand, it must not be assumed that the longer the acquaintance the better, because a number of factors enter after long acquaintance to introduce error in the results. A study that bears directly on this point was made with ratings of over 1000 public school teachers [14]. The most obvious tendency was to overrate persons who had been known longer. In "general efficiency," of those known less than one year only 10 per cent were rated excellent, of those known from one to 7 years 47 per cent were rated excellent, and of those known from 8 to 25 years 68 per cent were excellent. One possible explanation is, of course, that those who had been known many years actually had been teaching many years and had improved in efficiency as a result. However, other studies have shown that skill in teaching does not improve with experience to anything like the extent required to explain these results. Moreover, when the teachers are rated as to "physical efficiency" much the same trend is found, and it is scarcely plausible that physical efficiency should improve in this fashion with age.

The results can be explained satisfactorily on the basis of the acquaintance factor. A supervisor would dislike to concede that the persons under him had not improved under his supervision and if he rated them on a par with the more recent ones this would be tantamount to such a concession. Again one is apt unconsciously to identify himself with the older subordinates because they are more similar to him in age and this will result in more favorable consideration for them. Furthermore, one's own interests are apt to bias him in such identification. One supervisor who had previously been an athletic director gave as a reason for selecting a certain man as his best teacher the fact that he was a "he-man." Another supervisor who was a vigorous

Sunday school teacher selected a certain woman as her first choice because "she holds up high ideals before her pupils." Finally, with older subordinates, one becomes adapted to them and to some of their weak points. Various mannerisms and personality defects cease to attract attention so that ratings after long acquaintance are liable to be too high.

While these results were obtained in rating school teachers, the same reasoning would apply to executives or others rating their subordinates. The hesitation to concede that older employees had not profited by training under one, unconscious identification of the older with oneself, and adaptation to their weak points would operate in industry to introduce a similar error in ratings. It appears that knowing the subordinate too long decreases the critical value of judgments regarding him.

A somewhat similar situation was found in another instance when, not length of acquaintance but degree of friendship was considered. A group of persons rated one another in a number of traits, and also as to their degree of friendship with the rater [22]. It developed that there was a tendency to overestimate the good traits of one's friends. The traits that were overestimated in this way were quickness, proficiency, memory, persistence, adaptability, leadership, and scholarship.

Still another significant aspect of acquaintance is the condition of that acquaintance, that is, the conditions under which the individual has been observed for the most part. A school supervisor who has seen teachers primarily in the classroom and is rating them on various personality traits might give quite a different answer from friends who had seen them on the golf links, in the swimming pool, or at Joe's Place. Much naturally depends upon the vocational situation with reference to which the ratings are to be used. If the interest is primarily in classroom behavior, then perhaps the ratings made by those who have contact with the persons only in the classroom would be satisfactory. In addition to the query, "How long have you known the applicant?" it would be well to add, "How well do you know the applicant?" and, "Under what circumstances have you known the applicant?"

Bias. Other factors besides long acquaintance may produce bias. Many of us have individual prejudices against certain types of physiognomy, voice, or race, or against one sex. This point

was mentioned in discussing the criterion and the difficulties in securing accurate overall ratings from supervisors. It was noted there that one rater might be prejudiced against members of one sex and rate them accordingly. Race prejudice is another possible factor. Some raters may even have a bias against red hair or high-pitched voices. These prejudices are frequently "conditioned" by some unfortunate experiences with a few people belonging to a certain category. If, for example, one has been insulted by a member of a particular racial group or by someone with a certain cast of physiognomy, he is liable to develop a conditioned response of an unfavorable character toward such persons; subsequently he will give a poor rating to anybody belonging to those categories.

It is difficult to detect such a tendency in rating data except by detailed analysis. If one goes over the ratings by a given judge in detail and picks out the low ones he may find that they all have something in common. Or if he selects cases in which the raters disagree markedly he may find some explanatory factor. Marked deviation by a single rater from the average of a group of raters suggests a bias, or else misunderstanding of the procedure or the attachment of undue significance to some minor aspect which he should not have considered in making the rating. Attention should be called to two comprehensive references that deal with problems incident to the rater himself and with general sources of error in rating procedure [8, 27].

TRAINING RATERS

One of the most important aspects of the rating scale procedure is the training of the persons who are to do the rating, regardless of what particular form of scale is to be used. In this preliminary training a number of points should be particularly stressed and effort should be made to impress them upon prospective raters.

Attitude. One of these is the attitude with which the rater approaches his task. This should be objective and impartial. He must rate his friends on the same basis as other subordinates with whom he has only a business contact. One has merely to listen to two women discussing the merits of their children to appreciate the danger of being partial in making estimates. No

effort should be made to cover up a person's weak points, for if they are brought to light proper adjustments are often possible. Conscious prejudice sometimes is involved, but of more frequent occurrence is an unintentional bias due to special affability of the person rated or to a single incident favorable or unfavorable in character. It is a trifle difficult to give a poor rating to a man who is the "life of the party" or a high rating to one who has insulted you. It is important to teach the rater to abstract from all such things—to hold the individual, as it were, at arm's length and estimate him objectively and impartially.

Basis for Rating. Consideration must also be given to the basis on which the rater is to make his judgments. It is advisable for him to base his estimate on actual rather than expected performance. The latter kind of estimate becomes more subjective and involves not only the rater's ability to estimate traits from what he observes, but also his ability to infer therefrom how the person will behave at some future time. This is manifestly more precarious. Moreover, he should compare the employees he is rating only with one another. Messengers obviously should not be compared with typists. The ratings should be made with reference to the particular kind of work that is involved or the special industrial situation under consideration. Initiative in golf and in the cost department may be entirely different things. A man may be energetic in collecting stamps but lazy in figuring time slips. Patience in watching a cut with a machine tool does not necessarily reflect patience with one's family, and vice versa. Hence the rater should be taught to consider the traits of the man on the job rather than at home or elsewhere.

Standards. The rater obviously has to judge according to some standard, whatever the particular technique used. As previously mentioned, some may adopt a standard that is too lenient and others one that is too severe. This may usually be ascertained from a distribution curve of the ratings made by a given man. If he places most people too high or too low this should be pointed out to him in conference and he should be required to justify certain cases if he still maintains that his estimate is correct. He should be told at the outset that the persons below and above average are usually fewer in number than are average persons because they constitute exceptions to the general rule.

Frequently when a rater's tendency to overrate or underrate is brought to his attention he will revise his ratings and henceforth use a more normal standard.

Once a standard has been adopted by a rater he should make every effort to maintain it constantly throughout the procedure. There is danger of relaxing or otherwise changing the standard in the course of time. The man-to-man scale was devised in the light of this fact. With other types of scales the same standard can be maintained throughout after adequate training and practice. It is often well to recur occasionally to some of the ratings made earlier and see if they still seem correct. If they do, this indicates that the same subjective standard is being maintained.

Effort should be made, as described earlier, to distribute the ratings over a normal range rather than to bunch them. Some raters are afraid of making invidious distinctions and as a result give almost the same ratings to everyone. This should of course be called to their attention and they should be taught to distribute their ratings more widely. Another common tendency is to use greater care in making distinctions at the lower end of the scale than at the upper. Some raters bestow the better estimates rather indiscriminately, although they take plenty of pains with the poorer ones. The fine distinctions are often just as important vocationally at the upper end for determining promotional material as at the lower end for detecting misfits, and the rater should learn to govern himself accordingly.

Process of Rating. The essential aspects of the actual process of rating have already been brought out, but the rater should be watched to insure that he forms the habit of observing them. The ratings must be made independently. It is a temptation to talk them over with others who are making similar ratings. If a colleague glances at one's ratings and makes a casual remark, one is tempted to reconsider and perhaps to make some compromise. If the colleague is to be involved, the proper thing is for him to make similar ratings independently and then to compare the results statistically. It has been shown in various connections that greater validity is obtained by averaging independent estimates than by having the judges sit together as a committee and make a joint estimate.

The other aspect of the process of rating that is essential to

the success of most scales is judging one trait at a time. It is a temptation for the rater to take one individual and consider him throughout. This process is often the more expeditious. He should, however, be shown the danger of the halo effect and convinced of the desirability of employing the other method.

Sufficient Time. It is especially essential in training raters to convince them of the necessity of taking plenty of time. A busy executive who is accustomed to make quick decisions regarding matters of routine often finds it difficult or unpleasant to slow down and give the careful consideration to particular traits that is necessary. Consequently he must be "sold" on the value of the whole procedure so that, whatever the amount of time necessary for rating and rerating, he will be willing to devote that amount to the project. The point was made above that ratings of undesirable traits require a little more time than ratings of desirable ones.

Conference. Finally, to safeguard the whole procedure frequent conferences should be held between the one in charge of the project and the persons making the ratings. It is insufficient to give the raters printed directions and blanks and turn them loose. After they have had an opportunity to study the manual of directions it is a good plan to have a conference of all the men and talk it over. Any difficulties that have occurred to them can be clarified on the spot. Many of the points mentioned above in this section can be explained to them and emphasized, although subsequent repetition will of course be necessary. After this each one may well be asked to make out a sample set of ratings. These can be reviewed carefully and criticized in the light of the foregoing considerations. Ratings by different men may also be compared to advantage to find those who agree and those whose ratings seem typical. When any shortcomings in a man's ratings appear, his attention should be called to the fact. He can then rerate the same group or make other new ratings to see if he can profit by his previous mistakes. His second series of ratings may be similarly criticized and analyzed and perhaps compared with the first set, and this procedure repeated as often as necessary. In a large banking organization each rater has his ratings reviewed in personal conference three successive

times and this procedure is repeated twice a year if necessary [13].

The experience of a public utility is interesting in this connection. They were having difficulty in securing reliable ratings upon their service workers. Consequently they developed a system of interviewing the rater himself and asking various questions about persons whom he rated, with a view to checking the ratings themselves and also giving him some incidental instruction and practice in the rating procedure. They finally had 15 questions which were standardized for this purpose. A few of these will be mentioned:

1. "What are the good points of _____ as you see them?" This tended to bring out the actual basis on which the judgment was based and helped to run down any halo effects.

2. "Would you hire this employee again if you were to make the decision?" If the answer was "no" an explanation was requested.

3. "Does this employee have to be corrected or watched because of some special weakness?" This item sometimes brought out disciplinary contacts with the employee.

4. "Does this employee at times argue too much?" This tended to discover the cases which created an unfavorable impression because of arguments, with resulting halo effect [26].

This training of the rater tends to make his results more reliable. This has been shown statistically, as for instance in experiments with the officer's rating scale when a group of officers after training provided better estimates of intelligence than they did before instruction. As previously mentioned, the combined results of several raters are usually better than the results of one. A minimum of three independent ratings has been recommended on the basis of statistical studies. If, then, the rating scale has been properly constructed, if the raters have received adequate training, and if at least three raters make their estimates independently and the results are pooled, the results will be found of value in many practical situations.

SUMMARY

Rating scales are necessary in evaluating various traits that are of vocational significance but cannot be measured objec-

tively. Ratings made by interviewers, by previous employers or acquaintances, are used with a view to initial employment, and those by executives and foremen with a view to promotion or transfer. They afford a more uniform method of expressing opinion regarding prospective or present employees as they deal less with general impression or prejudice and more with specific traits. They educate the rater in leading him to make closer observations of his subordinates and in keeping the notion of personality before him, and they educate the employee who is rated in observing himself more critically. They often provide a valuable check on the progress of employees, and if they are on file they afford data to meet emergencies such as could not be obtained in systematic and reliable form on short notice.

In selecting the traits to embody in a rating scale for a particular situation it is desirable to eliminate those that are merely present or absent and not present in varying degrees. The best traits may be determined by circulating a questionnaire to persons familiar with the occupation, asking them to indicate those which they consider most important. The most frequently indicated traits may be included in the scale. A better procedure is to determine the traits in an interview or conference where ambiguities in terminology can be cleared up.

The next step is to weight the traits according to their relative importance. The frequency with which a trait is mentioned in the questionnaire or interview gives some idea as to its importance. The final list may be resubmitted to the executives with the request that they distribute a certain number of points among the traits; the average value assigned any trait may be taken as its approximate weight. In some cases the more reliable traits have been assigned greater weights not because the estimates are more closely related to the criterion but because they are truer indications of the trait under consideration. The weighting may be actually incorporated in the rating blank, or all the traits may be rated on the same basis and the weighting done subsequently.

It is necessary to define the traits in order to prevent the rater from putting his individual interpretation on a term. It is better to define in objective than in subjective terms because objective estimates have greater reliability than subjective. It is also de-

sirable to formulate the definitions with reference to the particular situation in which the scale is to be used.

The man-to-man rating method involves the construction of a master scale for each trait. This consists of the names of individuals who possess the trait in question in various degrees. Their names are written on the blank opposite appropriate rating values that have been previously determined. The persons being rated are compared man to man with the individuals on the master scale and given a rating similar to the number assigned the man on the scale whom they most resemble. The method was developed originally for military use but has been adapted to rating various occupational groups such as executives.

Another method involves rating the individual relative to other members of a defined group. The rater may be required to imagine that all the persons he knows who are engaged in the occupation in question are divided into five classes of equal ability and to locate the given individual with reference to these five classes. The blank may be presented in the form of a linear scale with the groups indicated by columns so that the rater can judge as finely as he wishes. A cruder scheme involves merely assigning each individual a particular number from 1 to 5, these numbers having been previously defined.

The graphic rating scale involves the name and definition of a trait or a question embodying it, followed by a line along which the rater checks at some point. He is guided by descriptive adjectives or phrases distributed along the line ranging from low degree of the trait to high degree. Care must be exercised in the selection of these adjectives or phrases to insure that the extreme ones are actually opposite and that the intermediate ones conform to the extremes. They should be spaced in accordance with the actual distribution of the trait and should perhaps be staggered with the high degree sometimes at the right and sometimes at the left, lest the rater drop into the error of making all his marks in about the same position. Graphic scales have been devised for many workers such as executives, secretaries, clerical workers, and salesmen. The ratings can be quantified by measuring the distance of the check mark from one edge or by using a stencil ruled in columns.

Any of the foregoing types of scales may be arranged with all

the traits on one page and a separate page for each subject, or with one trait and all the subjects on a page and a separate page for each trait. The latter procedure is more cumbersome for the rater but actually is preferable because it minimizes the tendency for general impression to influence all the ratings by a given person.

Another technique employs a check list. This may embody phrases for each trait like those in a graphic scale, except that they are listed in a column without any linear arrangement. Or a considerable list of miscellaneous statements may be provided, the rater checking them as plus or minus. These statements range from those indicating a favorable amount of whatever constitutes a good employee to those indicating an unfavorable amount. Scale values are determined for these items by the judgment of experts and by statistical analysis of estimates involving these items.

The reliability of a rating scale should be investigated before it is put into any very general use. Some notion of its reliability may be obtained by determining whether the ratings made by a person conform roughly to a normal distribution curve. If the curve is skewed toward the high or low end, or is very steep and narrow, it indicates that the rater is setting too strict or too lenient a standard or that he is failing to consider the whole range of the trait. It is often necessary to correct the original ratings in the light of this fact and to consider as high only those rated relatively high, and vice versa. Reliability may be further studied by noting the agreement of raters with each other. With the man-to-man scale there was a rather small agreement for different raters in the only systematic investigation reported. These discrepancies appeared to a considerable extent to be due to the construction of the master scales. With the graphic scale more encouraging results have been found. Different foremen rating the same subordinates agreed fairly closely in most instances. A further indication of reliability is given by comparing successive ratings by the same man. With the graphic scale rather high correlations were found between foremen's first and second ratings of the same men, and higher correlations still between their second and third ratings. One check-list scale showed still higher reliability.

The validity of ratings should be ascertained where possible, but often no criterion is available whereby to determine it. Estimates of intelligence in the Army scale showed some relation to intelligence as measured by a test, especially if the ratings by three or more judges were averaged. Certain items in a graphic rating scale for salesmen made some differentiation between those in different salary groups.

Certain sources of error in rating procedure may be noted. Traits that are subjective in character have appreciably less reliability than those that are more objective and that yield some products by which they may be judged. The halo effect is a particularly insidious source of error. This is the tendency to have a general impression of the individual and to rate him accordingly in all traits rather than to discriminate among the separate traits. It can be shown in many instances that estimates of different traits intercorrelate more highly than they ought to. The length of acquaintance with the person who is rated is of interest. If it has been long the rater is apt to give too favorable an estimate because of an unconscious identification of the older subordinates with himself, hesitation to concede that long exposure to his influence has not improved them, and adaptation to their weak points. The degree of friendship and the conditions of observation should also be considered. Bias and prejudice may play a role.

Finally, the raters ought to receive systematic training. They must be taught to take an impersonal, impartial attitude, and to rate the subordinate on actual rather than on expected performance and on performance in the special industrial situation under consideration. They must adopt a normal rather than an extreme standard as a basis for judgment and must maintain it throughout. The actual process of rating should be carried through independently and one trait at a time should preferably be considered for the entire group. The rater must be convinced of the importance of devoting ample time to the project. To safeguard the whole procedure, frequent conferences should be held to review the ratings with those who made them and to discuss any errors that are manifest.

If the rating scale has been properly made and at least three trained raters make independent judgments of a group of in-

dividuals, the combined results will be of some value in the practical situation.

REFERENCES

1. Bills, M. A. A Method for Classifying the Jobs and Rating the Efficiency of Clerical Workers. *Journal of Personnel Research*, 1923, 1, 384-393.
2. Bingham, W. V. Halo Invalid and Valid. *Journal of Applied Psychology*, 1939, 23, 221-228.
3. Bradshaw, F. F. Revising Rating Techniques. *Personnel Journal*, 1931, 10, 232-245.
4. Cook, H. E., and Manson, G. E. Abilities Necessary in Effective Retail Selling and a Method of Evaluating Them. *Journal of Personnel Research*, 1926, 5, 74-82.
5. Dorcus, R. M. Some Factors Involved in Judging Personal Characteristics. *Journal of Applied Psychology*, 1926, 10, 502-518.
6. Freyd, M. A Graphic Rating Scale for Teachers. *Journal of Educational Research*, 1923, 8, 433-439.
7. Gallup, G. A. Traits of Successful Retail Sales People. *Journal of Personnel Research*, 1926, 4, 474-482.
8. Guilford, J. P. *Psychometric Methods*. New York, McGraw-Hill, 1936, 566 pp.
9. Hollingworth, H. L. *Judging Human Character*. New York, Appleton-Century, 1923, 268 pp.
10. Jones, E. S. Personality Terms Commonly Used in Recommendations. *Journal of Personnel Research*, 1924, 2, 421-430.
11. Kenagy, H. G., and Yoakum, C. E. *The Selection and Training of Salesmen*. New York, McGraw-Hill, 1925, 380 pp.
12. Kingsbury, F. A. Making Rating Scales Work. *Journal of Personnel Research*, 1925, 4, 1-6.
13. Kingsbury, F. A. The Principles Involved in Securing Service Ratings as Exemplified in a Large Bank. *Public Personnel Studies*, 1925, 3, 70-84.
14. Knight, F. B. The Effect of the Acquaintance Factor upon Personal Judgments. *Journal of Educational Psychology*, 1923, 14, 129-142.
15. Kornhauser, A. W. A Comparison of Raters. *Journal of Personnel Research*, 1927, 5, 338-344.
16. Laird, D. A. *The Psychology of Selecting Men*. New York, McGraw-Hill, 1925, 274 pp.
17. Miner, J. B. The Evaluation of a Method for Finely Graduated

- Estimates of Abilities. *Journal of Applied Psychology*, 1917, 1, 123-133.
18. Paterson, D. G. The Graphic Rating Scale. *Journal of Personnel Research*, 1922, 1, 361-376.
 19. Richardson, M. W., and Kuder, G. F. Making a Rating Scale That Measures. *Personnel Journal*, 1933, 12, 36-40.
 20. Rugg, H. O. Is the Rating of Human Character Practicable? *Journal of Educational Psychology*, 1921, 12, 425-438, 485-501; 1922, 13, 30-42, 81-93.
 21. Scott, W. D., and Clothier, R. C. *Personnel Management*. Chicago, Shaw, 1923, 643 pp.
 22. Shen, E. The Validity of Self Estimates. *Journal of Educational Psychology*, 1925, 16, 104-107.
 23. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
 24. Symonds, P. M. Notes on Rating. *Journal of Applied Psychology*, 1925, 9, 188-195.
 25. Thorndike, E. L. A Constant Error in Psychological Rating. *Journal of Applied Psychology*, 1920, 4, 25-29.
 26. Wadsworth, G. W. Fit Employees to Their Jobs. *Personnel Journal*, 1937, 16, 165-170.
 27. Weiss, L. A. Rating Scales. *Psychological Bulletin*, 1933, 30, 185-208.

Chapter XIII

MISCELLANEOUS DETERMINANTS OF VOCATIONAL APTITUDE



VALUE

Employment psychologists have devoted most of their efforts to the use of mental tests of one sort or another for the prediction of vocational aptitude. This is due considerably to the fact that the tests are objective and yield results that do not depend on the judgment of the applicant or of persons familiar with him. The tests, moreover, are quantitative and usually yield a wide range of scores. All these things contribute to the reliability and validity of the results.

Supplement Tests. Granted that test procedure is generally superior to less quantitative or objective methods, there is nevertheless the possibility that these latter may be valuable as a supplement to the tests or even in lieu of them in instances where tests are not feasible. With reference to the former possibility we have previously seen that, in deriving a regression equation for predicting vocational aptitude, the more variables evaluated, the greater the chance of finding a group which, if properly weighted, will give a high correlation with the criterion. For the average marksman a shotgun is more effective than a rifle. So with a group of tests or other measurements none of which can give a perfect vocational prediction, the more that are tried, the greater the chance of finding some that are valuable for the purpose at hand.

In the discussion of weighting a group of vocational tests, it was suggested that it is advisable to try out a rather wide range of tests and select for further careful study those which have high correlations with the criterion and low correlations with

each other. It often develops in an employment research that most of the tests used intercorrelate rather highly. Hence there is the possibility of turning to other variables besides tests—e.g., such things as items of personal history—that may perhaps show some correlation with the criterion and likewise a low correlation with the tests. If such unique variables are found, their addition will increase appreciably the validity of the whole procedure of prediction. At any rate, it has seemed worth while in many instances to determine whether any additional variables of this sort are available and to evaluate them at least in a rough statistical fashion with a view to further refinement of treatment, providing they are promising. It is quite possible that tests plus certain miscellaneous factors will give a better prediction of occupational aptitude than will tests alone.

In Lieu of Tests. In some employment situations it is not feasible to embark on a scientific testing program with a view to developing employment techniques. Perhaps the concern cannot at the time afford the necessary outlay or it is inadvisable to take the employees away from their work long enough to test them. Perhaps the present number of workers is too small for statistical purposes, but records of a biographical nature and production figures are available for a larger number of former employees. In such cases some of these miscellaneous factors may be used in lieu of tests and prove better than nothing. Moreover, various methods are ordinarily in unsystematic use, such as letters of application, recommendations, and interviews, which can be systematized to advantage or be evaluated statistically to determine whether they are actually worth using. The following factors will be discussed in the present chapter: academic record, initial success in the vocation, personal history blank, letter of application, recommendations, and the interview.

ACADEMIC RECORD

It is often a simple matter for the employer to obtain a transcript of the applicant's academic record in school or other educational institution. Many application blanks call for the grade finished in school. But while this may give a rough indication of educational attainment, it is doubtless better to obtain school marks or something analogous. Where the situation warrants,

it is often possible to write to the institution which the applicant attended and obtain information regarding his educational career. This practice is especially common in the case of persons who have attended technical institutions and apply for positions along the technical lines pursued.

School Progress a Selective Procedure. There are *a priori* grounds for believing that school progress should give some indication of subsequent success. The school itself has probably exercised a certain amount of selection among its pupils. Some individuals are able to meet the normal educational demands and progress at the ordinary rate. Others, however, are unable to meet these demands; they fall behind or perhaps drop out rather early in their educational career. Still others, on the contrary, may be able to progress more rapidly because of their superior capacity. Thus indirectly the rate of progress in school, especially with reference to advancement or retardation, gives some indication of capacity to meet the problems and demands of the school situation.

Similar principles apply to the grades or marks received in school. These should in the long run reflect the student's actual accomplishment and this in turn give some indication of his ability. These suggestions must be qualified in the light of the fact that students do not always use the ability they possess and hence their grades may be an unreliable indication of that ability. Moreover, if a school system is poorly organized and has inadequate methods of grading or promotion, little significance can be attached to the results. However, in the general case there is some ground for the assumption that the school curriculum is after all a rather prolonged mental test.

Early Academic Record Prognostic of the Later Record. Various statistical studies have been made to determine the prognostic value of school marks. For instance, it has been shown that grades obtained early in the academic career are quite indicative of marks obtained later therein. Many data of this sort are available, but detailed presentation is not warranted. Marks in the 7th grade correlate with those in high school to the extent of .72 [6, 177]. About 70 per cent of those in the upper half of their class in high school are in the upper half in college. If a student is in the upper quarter in high school the chances

are about four out of five that he will be in the upper half in the university. Of Harvard graduates who entered law school with a plain degree—i.e., with no distinction—only 7 per cent received a degree with distinction in law, while for those graduating *summa cum laude*—i.e., highest distinction—the corresponding figure was 60 [16].

Academic Record and Occupational Success. The more important problem from the employment point of view is the extent to which school marks may be indicative of subsequent proficiency in industrial or professional activities. A study of graduates of Wesleyan University throws some light on this problem [20]. The students who graduated between 1860 and 1889 were divided into three groups—those who graduated with valedictory or salutatory honors, i.e., ranked either first or second among their graduating classmates in scholarship; those who were elected to Phi Beta Kappa, an honorary fraternity for which high scholarship is the prerequisite; and those who achieved no such honors. The percentage of each group appearing in the 1914 edition of *Who's Who* was then computed. These percentages are given in Table 44. It is obvious that the honor men and

TABLE 44. PERCENTAGE OF COLLEGE GRADUATES
FOUND IN "WHO'S WHO"¹

Group	Per Cent
Honor men.....	48
Phi Beta Kappa.....	31
Others.....	10

the members of Phi Beta Kappa stand a much higher chance of distinction of the type under consideration. The group which took no academic honors or distinction constituted about two-thirds of the entire group, but actually contributed only about one-third of the graduates who appear in *Who's Who*. To be sure, the type of success that lands one in *Who's Who* is apt to be literary, professional, political, or academic, rather than industrial or commercial. Unfortunately, an analogous criterion is not available for these latter types of success.

Another study was made of 240 alumni of this same institu-

¹ After Nicholson.

tion [15]. The alumni secretaries estimated them as to success in three degrees—successful, average, unsuccessful—and these degrees were arbitrarily scored as 5, 3, and 1 respectively. An activity score was also derived based on their participation as students in extracurricular activities. A high and a low scholarship group were selected; the average success scores for these two groups were respectively 4.35 and 2.91. The critical ratio for the difference between these averages is not included in the published account, but from the variability figures a rough approximation indicates that the difference is clearly significant, with a critical ratio of probably 8 or 10. From another angle, 89 per cent of the good-scholarship group exceed the median success score of the low-scholarship group. Incidentally, the results are equally striking with reference to participation in extracurricular activities. Students making high ratings in this respect had an average success score of 3.56 and those making a low rating had a success score of 2.3.

A study was made of the graduates of a technical institute in mechanical and electrical engineering, comparing marks at the institute with subsequent salary. Men in the graduating classes of three successive years were studied and their salaries obtained from four to six years after graduation. While success in engineering vocations may not be entirely reflected in salary and other factors besides proficiency may influence salary, nevertheless it gives some indication of vocational success. The men were divided into four groups on the basis of their school marks and the average salary obtained by each group was computed. Table 45 shows the results [6, 198]. To facilitate comparison, the salary of the highest group is taken as 100 per cent in the last column and the others reduced to percentages thereof. It is obvious that the men who had better records while at the institute obtained appreciably higher salaries on the average. (These salaries were for 1913.) The differences are not large, but they indicate a trend. If the individual salaries are correlated with individual marks, the coefficients for the graduates of each year are all positive and average .27. This correlation would not warrant academic record being used as the sole means of predicting vocational aptitude, but such a record might prove of

TABLE 45. SCHOLARSHIP IN A TECHNICAL INSTITUTE AS INDICATIVE OF SUBSEQUENT SALARY²

Scholarship Group	Average Salary	Per Cent Salary
Highest quarter.....	\$1664	100
Second quarter.....	\$1462	88
Third quarter.....	\$1418	85
Lowest quarter.....	\$1279	77

some value, as above suggested, in supplementing other indications, especially if it was not highly correlated with the other variables.

The academic records of over 4000 graduates of West Point from 1818 to 1905 were studied with reference to subsequent success [25]. The criterion of success was taken as appointment to the rank of brigadier general or above. Whereas 29 per cent of those in the highest fourth of their class in scholarship achieved the rank of brigadier general, only 15 per cent of those in the lowest quarter did so. The results are more striking when we consider only the men at the extremes of their graduating classes in scholarship. Of all the men who ranked highest in their graduating class, 47 per cent were successful, but for those who ranked lowest the figure was only 6 per cent.

A study of employees in the Bell Telephone System related their scholarship standing in college to their subsequent salary in the organization [4]. The essential features of the relationship are shown in Fig. 8. The salaries listed on the ordinate are reduced to terms of the median or average salary for the entire group. The abscissa represents the number of years since graduation. The different curves are for the different scholarship levels as indicated. For example, the students in the highest 10th of their graduating class were receiving salaries fifteen years later on the average 20 per cent above the average salary of the entire group under investigation. After thirty years their differential was about 55 per cent. Those in the lowest third of their

² From H. L. Hollingworth, *Vocational Psychology*, by permission of D. Appleton-Century Company, New York.

classes obviously went down about 20 per cent through the thirty-year period.

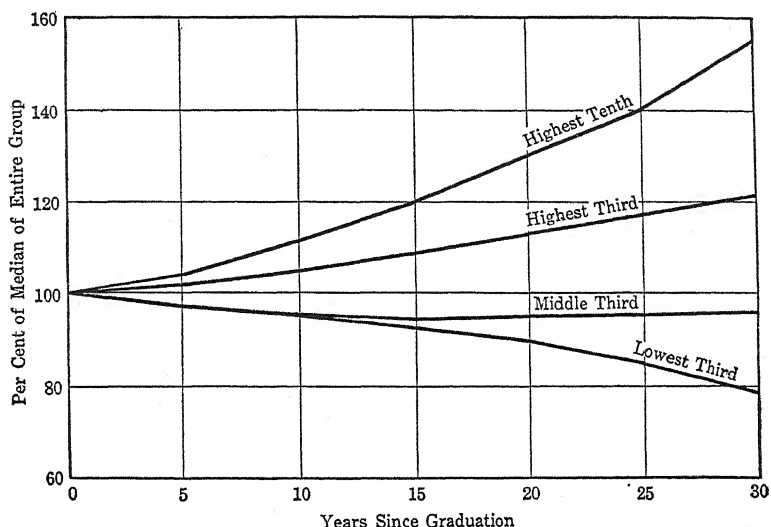


FIG. 8. SALARY AND COLLEGE SCHOLARSHIP RANK

A similar, though less marked, trend is found when campus achievement in extracurricular activities is considered in relation to subsequent salary. Of students who were in the first third of their class in scholarship, had made substantial campus achievement, and had earned over two-thirds of their college expenses, 68 per cent were in the top third in salary.

Amount of Education. It is common practice when obtaining information from an applicant to ask what grade in school he finished, in place of a transcript of grades received. This has some significance. In many cases the employer is interested in whether the applicant has certain educational fundamentals which will be actually necessary for his work. He may need a certain amount of arithmetic, such as fractions, in order to make out time slips or compute dimensions of material that is to be used. He may need a certain proficiency in reading in order to interpret typewritten directions or orders that are issued. If he has not progressed beyond a certain grade in school, it is prob-

able that he has not been exposed to fractions or to reading of the requisite difficulty.

Another aspect of the matter is significant with the younger generation. In these days of compulsory education, the grade finished in school is an indirect indication of intelligence. Suppose that in a given state everyone is compelled to attend school until the age of sixteen. If, then, one individual has finished the third year of high school and another only the seventh grade, both having attended school some eleven years, it is obvious that the latter has occasionally failed to be promoted. This may indicate poor teaching or improper motivation by parents and others, but it also probably indicates a difference in the innate intellectual capacity of the two persons. The same information may of course be obtained, if the applicant's statement can be trusted, by asking him both the grade completed before leaving school and his age at leaving. The tendency for pupils of high intelligence to progress more rapidly in school when given opportunity has been repeatedly demonstrated. Hence, rapid progress may give some presumption of greater intellectual capacity. In situations where tests are not used, some inkling as to the applicant's intelligence may be obtained in this indirect fashion.

Comparison of years of schooling with occupational criteria yields results which are suggestive but not very striking. With a group of billing-machine operators the number of years of schooling gave a correlation with speed in billing of .23, and with accuracy in billing of .31 [13]. For clerks in an insurance company the correlation of years of schooling with grade of work was .47 [35]. With students of telegraphy no correlation was found between years of schooling and receiving ability after 100 hours of practice [34].

An adding-machine company found that 50 per cent of its superior salesmen were college men, 30 per cent had attended high school or business school, while 20 per cent had only a grade school education. However, when all the men of the sales force were considered, only 12 per cent of the college men were "A" salesmen, whereas 20 per cent of the grade school men were in this class [10, 224].

In another concern 45 per cent of the successful salesmen were college men and only 35 per cent of the failures were col-

lege men. Another company found that men with high school education made more successful salesmen than those with more or less than this amount of schooling. This seemed true in some insurance companies; but in another group of insurance men there was a correlation of only .11 between years of schooling and production, and the college men seemed best, grade school men the next best, and high school graduates the worst.

Findings like the foregoing point to the necessity of evaluating a particular variable, such as education, with reference to the particular situation in which the variable is to be used. Such a factor may be of some value for vocational prognosis in one organization and worthless in another.

Academic Record in Special Subjects. Although general educational attainments give indirect evidence regarding intellectual capacity, there is a further possibility that effort or achievement in special educational subjects may afford some indication of special capacity or interest that will be of vocational significance. The vocational implication of some of the more extreme cases is obvious. A person who has shown aptitude for mathematics by achieving good grades in his mathematics courses will qualify, other things being equal, for industrial work in which it is necessary for him to make computations. Similarly, a man who, according to school records, has done well in manual training has thereby demonstrated some mechanical proficiency, and the expectation of his being successful in mechanical work is consequently somewhat greater.

If an individual has had the opportunity to elect certain school subjects rather than others, his choices may reflect either his interest or his ability, or both. The average pupil selects school subjects which he likes and usually those in which he is fairly proficient. Persons who, for instance, have voluntarily chosen to study mathematics or natural science will perhaps stand a better chance in engineering occupations than will students who of their own choice pursued history or the classics. In this connection, however, it is essential to determine whether the choice was the applicant's own or whether it was the result of influence by relatives or friends. While it has been discovered that some of the most successful engineers had a classical education, this merely reflects the fact that their parents had been of a

high order of intelligence, had consequently obtained a liberal education in the day when only the more intellectual went to college, and had then encouraged their children to pursue the same type of classical education. These children, inheriting the high intellectual capacities of their parents, were destined for reasonable success in almost any line they might pursue.

Inferiority of Academic Record to Actual Tests. These aspects of the academic record should not be used to the exclusion of the quantitative measurements which have been described in earlier chapters, unless absolutely necessary. It can be shown that these factors are not as valid as are mental measurements in predicting occupational success. Even if school records are available in quantitative form so that it is unnecessary to take the applicant's word as to his educational career, these records have been shown to be less satisfactory than mental tests.

A case in point is the relative validity of high school marks and of specific tests in predicting success during the first two years in an engineering college [36]. Table 46 gives these cor-

TABLE 46. CORRELATION OF FIRST TWO YEARS' WORK IN ENGINEERING COLLEGE WITH HIGH SCHOOL GRADES AND WITH SPECIAL TESTS³

High School Grades		S.P.E.E. Test Scores	
Algebra.....	.21	Arithmetic.....	.45
English.....	.21	Algebra.....	.30
Geometry.....	.22	Geometry.....	.35
Physics.....	.25	Intelligence.....	.29
Chemistry.....	.29	Physics.....	.36
		Technical information.....	.22
<hr/>		<hr/>	
Total.....	.28	Total.....	.48

relations. The left part of the table shows the correlations between high school grades and the criterion. Algebra is the least predictive and chemistry the most, with the validity for the total record only .28. The same students were given the test for engineering aptitude devised by the Society for the Promotion of Engineering Education. This test comprises six parts, each occupying about thirty minutes. The validities of these six parts appear in the right part of the table. It is to be noted that the

³ After Thurstone.

thirty-minute test for arithmetic gives the best prediction of any single measure. It is also to be noted that in every instance a thirty-minute carefully standardized test dealing with specific information in a school subject is more predictive than the entire high school record in that particular subject. High school grades in algebra, for instance, correlate .21 with college grades, while the special algebra test correlates .30. The correlation of the total test score with college work is .48. This may be contrasted with the corresponding correlation of .28 for the high school grades. Thus school grades are inferior to actual scientific measures for vocational prediction. They should be used in place of the latter only when it is impossible to obtain the psychological measures. Whether or not school grades are valuable in supplementing such measures must be determined in the particular vocational situation.

INITIAL SUCCESS IN THE SAME OCCUPATION OR A RELATED ONE

In some cases it is possible, if a person's production record in a given occupation over a short period of time is known, to predict his subsequent efficiency. An investigation of this sort was conducted with insurance salesmen [5]. The first group was small, but records were available for four years. Group II was larger and three years' records were available. Group III was larger still, but only two years' records were available. With these data success the first year can be compared with success in subsequent years. The correlations between production in different years are shown in Table 47.

TABLE 47. CORRELATIONS BETWEEN SALES PRODUCTION IN DIFFERENT YEARS⁴

	Group I	Group II	Group III
First year and subsequent year92	.72	.74
First year and second subsequent year76	.68	
First year and third subsequent year47		
First year and total production90	.88	.93

⁴ After Goldsmith.

It can be seen, for instance, that with Group I the correlation between first-year production and production the subsequent year is .92, whereas the correlation between the first year and the second subsequent year is .76, and that between the first year and the third subsequent year is .47. This same tendency is indicated in Group II, namely, the first year gives a better indication of the first subsequent year than of the later years. The figures at the bottom of the table which indicate the correlation of the first year and total production of all the years are quite large.

The number of accidents is sometimes taken as an inverse indication of a worker's efficiency. In a machine shop the correlation of the number of accidents in successive quarters of the year was computed, i.e., the tendency for a worker to have in a given quarter the same number of accidents as he had in the preceding quarter. Four such correlations for successive quarters are .72, .37, .53, .69 [14, 218]. A worker with a record of accidents is more liable to have others than is a worker with a clear record. He apparently does not profit from the experience. The accidents seem to be due to some fundamental cause. In so far as they are an index of inefficiency, early failings in this respect are prognostic of later ones.

Among billing-machine operators efficiency from the standpoint of both speed and accuracy in the sixth month of work was studied to see how well it could be predicted from earlier efficiency and also how well it would predict later efficiency [13]. The correlations are shown in Table 48. The work during the first month is worthless from a predictive standpoint. From then on it appears somewhat diagnostic. This is especially the case with speed, which becomes of some significance in the second or third month. Accuracy has little predictive value until almost the fourth month. The correlations of the sixth month with adjacent months are of course higher than with more distant months.

That it is advisable to investigate the predictive value of early success on the job for the individual operation rather than for the entire industry is brought out in the following study [3]. In a knitting mill typical practice curves were obtained in three of the operations week by week for about a year. Production for

TABLE 48. CORRELATION OF EFFICIENCY AT A BILLING MACHINE DURING THE SIXTH MONTH OF SERVICE WITH EFFICIENCY DURING OTHER MONTHS⁵

	Speed	Accuracy
First month.....	.20	.00
Second month.....	.48	.25
Third month.....	.60	.31
Fourth month.....	.76	.58
Fifth month.....	.76	.74
Seventh month.....	.86	.72
Eighth month.....	.74	.40
Ninth month.....	.68	.16

the first ten weeks was correlated with that for weeks 41 to 50. These correlations for a trimming operation were .61, for a covering operation only .01, and for hemming .27. Obviously, there was a tremendous variation with the different jobs; in some the early performance would be reasonably diagnostic of the later, and in others quite the contrary. In the former case, workers with poor initial records may well be considered for transfer to another job for which they are better adapted.

It is sometimes suggested that experience in some type of related work would be prognostic of success in a given type. This proposal raises the broad question of transfer of training—whether what one learns in one kind of work will be transferred to another kind. The classical example was the assertion that certain academic studies “trained the mind” so that a person would be more effective in almost any other type of career. This problem has been studied experimentally. Without going into details, the upshot of our thinking now is that transfer occurs only in so far as there are identical elements in the two situations. Practice in simple addition, for example, might transfer to multiplying two-place numbers by two-place numbers because in the latter operation after all some addition has to be performed. In the industrial situation the amount of transfer from one job to another would depend largely on how much they had in common. If they both involved using a hammer to hit something, skill with that tool in one job might be transferred to another. A mistake which is sometimes made, however, is to

⁵ After Kornhauser.

ascertain merely at what similar work a person has had previous experience, regardless of his actual efficiency in that work. With this procedure there is the danger of actually hiring people who were in the wrong job originally and perpetuating that maladjustment by putting them into a similar job where they will likewise be maladjusted.

PERSONAL HISTORY OR APPLICATION BLANK

A personal history or application blank is often filled out as a preliminary to an interview. This is desired sometimes in order to save the interviewer's time, and sometimes to sort out in a preliminary way from a group of applicants those who are worth interviewing. The blank aims to bring out the more obvious data regarding a person's capacities and interests and may form a basis for subsequently securing more detailed information. It may be filled out entirely or partially by either the applicant or the interviewer. It is sometimes arranged so that the applicant fills one side and the interviewer uses the reverse. At any rate, some such blank is found in most employment offices.

Technique of Evaluating Items in Blank. These personal history blanks are generally used rather uncritically. It is assumed, perhaps on the basis of casual observation, that certain items, such as age or marital status, are prognostic of occupational success. This assumption may not be as erroneous as the assumption that a bump on the head just above the ears indicates ability at constructing things or that fine-textured skin presages artistic achievement. But it is nevertheless an assumption, whereas science prefers to deal with facts. These various items of personal history can be evaluated statistically and the facts obtained. After a group of individuals has been on the job sufficiently long to demonstrate their ability, it is possible to determine whether certain items actually differentiate the good from the poor workers. If, for instance, a group of salesmen are divided into successes and failures, and it develops that most of the successful salesmen are married and most of the failures are single, this information as to marital status may be of some significance in employing salesmen in the future.

It is not feasible in this type of problem to employ any rigorous statistical analysis. About the best that can be done is to divide

the individuals into two classes as far as the criterion is concerned. These two classes may represent a division at the midpoint of the range of occupational ability, or preferably may consist of classes at the two extremes. After such a division is made, however, it is a simple matter to tabulate for any particular item on the history blank the percentage of the successful group giving or failing to give a certain answer, and similar percentages for the unsuccessful group. If a particular answer is given much more frequently by the successful individuals than by the unsuccessful, that item or answer may be taken as to some extent differential of success in the occupation in question. If there is any doubt as to whether the difference is large enough to be significant, recourse may be had to proper formulae for determining this significance. (Cf. p. 329.)

Physical Characteristics. The usual application blank calls for such items of a physical character as height and weight. Hence it is well to determine whether there is any relation between these and fitness for the particular job. Of course in certain cases it is perfectly obvious that large stature is desirable. In hauling a heavy truck or doing work where great force must be exerted, a large man has an obvious advantage. Such patent examples as this need no scientific study. There are, however, more subtle possibilities in stature. One that has sometimes been rather seriously considered by employers of salesmen is the possibility that large salesmen can "dominate" the prospect and hence make more sales. One manager actually attempted to develop a sales personnel over six feet in height. Some of us have occasionally felt a bit inferior in the presence of a large man with a dotted line. The writer makes it a practice when interviewed by a Gargantuan salesman to have the man seated, and if possible to sit on the desk himself, so as to dominate him rather than be dominated by him.

Some statistical evidence is available on this matter of stature. In two concerns salesmen were divided into three classes of approximately equal size on the basis of their sales records [12]. The average height and weight of each group are given in Table 49. Within these groups there is evidently little relation between stature and selling. In height the group with medium selling

TABLE 49. AVERAGE HEIGHT AND WEIGHT OF SALESMEN OF DIFFERENT DEGREES OF EFFICIENCY⁶

Sales Record	Company A		Company B	
	Average Height (inches)	Average Weight (pounds)	Average Height (inches)	Average Weight (pounds)
Highest third.....	69.0	156	69.3	180
Middle third.....	68.6	153	68.0	185
Lowest third.....	69.8	158	69.0	178

ability has the lowest average. The weights likewise are equivocal. In one company the poorest salesmen are slightly the heaviest, while in the other the medium group is most ponderous.

Results of this sort, however, do not always appear [10, 219]. An insurance company found that the average monthly sales of men under 69 inches in height were \$740, while the sales of those over 69 inches were \$1165. In another concern the average height of the ten leading salesmen was 70.7 inches, and the average height of all failures was 69 inches. In another group men weighing between 140 and 180 pounds were found to average higher in monthly production than those above or below those limits.

But this is not the whole story. While there may be no universal tendency within a given salesgroup for the larger men to be more effective, the evidence is clearer that salesmen as a whole are larger than the average individual. The results of a number of studies, including the one in the preceding table, are summarized in Table 50. The average height and weight are given for various sales groups. For comparison with the general population the average height and weight of about 1,000,000 men in the Army are included. The average height of some 220,000 men tabulated by the Association of Life Insurance Medical Directors is also given. This is somewhat greater than the Army average, but not as large as the average of any of the

⁶ After Kitson.

TABLE 50. AVERAGE^a HEIGHT AND WEIGHT OF DIFFERENT GROUPS OF SALESMEN⁷

	Average Height (inches)	Average Weight (pounds)
Mixed group.....	69.6	170
Insurance.....	69.5	
House-to-house.....	69.5	158
Technical.....	69.3	169
Miscellaneous A.....	69.1	155
Miscellaneous B.....	68.8	181
Miscellaneous C.....	69.5	160
General population—Army.....	67.5	142
General population—actuarial data.....	68.5	

^a For the smaller groups the median was used rather than the mean.

sales groups. These latter are considerably superior to the general population in both height and weight.

The difference should be somewhat qualified in the light of the fact that the Army group was somewhat younger than the others. Most of the groups of salesmen average in their thirties; the average man in the Army sample was well below this. Many persons, of course, put on weight as they grow older and the salesmen might have been heavier partly because of their maturity. It is rather doubtful if this would account for differences as large as most of those in the present case. Moreover, results for height would be much less affected by this error because this characteristic changes little after one reaches maturity.

While there may be some doubt regarding the relation of stature to production within a given sales organization, there is little question but that salesmen as a whole are larger than their prospects. If any conclusion other than this is to be drawn, it is that perhaps men of medium stature, although above that of the general population, are somewhat more effective than those at either extreme. It has been more or less seriously suggested that such a salesman is large enough to dominate his prospect effec-

⁷ After Kenagy and Yoakum, *et al.*

tively, but not too large to get around easily and cover the ground.

Age. Considerable significance is attached to age in employment and analogous problems. Some railroads will not employ a man who is over 35 and retire employees on a pension at the age of 65 or 70. The teaching profession in some cases has similar retirement rules. Some states will not permit a person under 16 to drive an automobile. A citizen must be 21 in order to vote. Some states set minimum age limits of 14 to 16, below which an individual cannot be employed in industry.

Such tendencies are usually based on popular belief that persons outside of the age limits in question are ineffective in the type of work under consideration. It is in point, then, to consider more systematically any psychological aspects of age that may be of vocational significance. We know, of course, that mental proficiency does change in one's early years; the changes at the other extreme are obvious. The influence of age on performance in certain mental tests was mentioned in Chapter VII. Proficiency in all the tests increased from childhood up into the teens, but the rate of increase was not uniform. Likewise at the other extreme, proficiency decreased considerably in some tests and much less in others. It is quite possible that rather extensive age differences of this sort exist, and if so, some of them may be of vocational significance.

An obvious approach to the problem from the practical standpoint is to correlate age with occupational proficiency and to determine within a particular group of employees if the more mature are the more proficient. With clerical workers in the civil service a correlation of .06 was found between age and efficiency scores. For another group of clerks there was a correlation of .35 between grade of work done and age [35]. With a group of telegraphers the correlation between age and receiving ability was $-.09$. Among insurance salesmen production correlated with age at the time of initial contract with the company to the extent of .15. Only one of these coefficients is large enough to be of any possible value. This does not tell the whole story, however, because it may be that persons of medium age are most efficient rather than the oldest ones, whereas a large correlation would not be obtained unless the oldest ones were the best.

This factor may be investigated by noting the relative efficiency of workers of different ages, with a view to determining whether there is an optimum age for a given occupation. For a miscellaneous group of superior salesmen the average age was almost 39. Only 11 per cent of them were under 30 and only 10 per cent over 50 [10, 217]. Those of middle age were manifestly the big producers. This, of course, suggests that the younger men had not had sufficient experience, and to some extent this is the case. In an insurance company where men with previous insurance experience were generally more efficient and where the best producers were between 35 and 50, it developed, nevertheless, that the best producers at the time of contract—many of them without previous experience—were between the ages of 30 and 45. Even apart from experience it seemed that maturity was desirable. Similar studies with other groups of salesmen have revealed the fact that extremes of age are somewhat less favorable than the middle range.

It might often be worth while with other kinds of occupations to apply similar techniques and determine whether there seems to be any optimal age at the time of initial employment. There are doubtless many types of work in which maturity is necessary in order to impress favorably persons with whom one deals, and there are other types in which a man who is too old will fail because of decreased mental efficiency. It is necessary to answer the question statistically in any given case.

Another aspect of age should be mentioned. Quite apart from efficiency, there is a possibility that age may bear some relation to stability or turnover. Several studies have been made of this relation, but rather than revealing any specific effect of age as such, they have brought out various other complicating factors that enter into different age groups. A case was mentioned previously (p. 174) in which a bonus system was ineffective with very young female employees because they took their pay envelopes home unopened and their parents received the bonus.

A study was made of the workmen who quit in two large firms, one doing metal work and the other manufacturing furniture [11]. These "quits" were classified as to age in five-year intervals and records were tabulated to show the average number of weeks worked by employees in a given group before they left.

Table 51 gives the results. Both companies show a manifest turnover among the younger workers. This doubtless reflects the

TABLE 51. AVERAGE NUMBER OF WEEKS WORKED BY
EMPLOYEES WHO QUIT⁸

Age	Average Number of Weeks	
	Company A	Company B
Under 21.....	18	10
21 to 25.....	19	9
26 to 30	23	11
31 to 35	31	24
36 to 40... ..	28	19
41 to 45.....	29	12
46 to 50.....	30	8
51 to 55.....	58	15
Over 55.....	56	25

natural instability of youth and the legitimate search for a vocational objective. At the other extreme there is marked stability for those older than 50. At this time one's interests have become fairly well established and profitable change in employment is rather unlikely. Likewise, between 30 and 35 there seems to be considerable stability, this being a period when many individuals buy homes or rear families. From then on until 50 there is something of a decrease. It is quite possible that at this period the worker's family is becoming more self-supporting and his domestic responsibilities are not quite so pressing. He realizes that old age will come soon and that he had better change his occupation now if at all. Consequently, he takes this opportunity to try other occupations with a view to finding one that will be permanent and satisfactory. Incidentally, the results suggest the desirability of watching for symptoms of unrest at these critical ages, being more tolerant of the workman, and attempting to make such adjustments as will keep him at the job if he is satisfactory.

⁸ After Kitson.

A study of workers' satisfaction as related to age is at variance with the foregoing, but the technique employed was somewhat different. Data were obtained from 273 men in hobby groups that met in connection with an adult education program. Among other things they filled out a questionnaire about their hobbies and also about their satisfaction with their present job. These data were broken down by age with the following results: At ages 20 to 24 the percentage expressing themselves as satisfied with their job was 72. At ages 25 to 34 the percentage was 48, an obvious drop. At 35 to 44 the percentage rose again to 75 and stayed up pretty well thereafter. The interpretation made by the investigator was that the youngest ones were glad to have almost any job. Those between 25 and 34 were a bit dissatisfied because they wanted to get ahead more rapidly, and then as greater achievement came in middle life this brought with it increased satisfaction. The difference between the trend in this study and in the one previously cited may be partially due to the actual time at which the study was conducted. The first one cited was made in 1922 and the second in 1938 and 1939. Unemployment was more widespread in the latter period and it seems plausible that satisfaction at having any kind of a job might have been more pronounced among the youth in the more recent period. This suggests a broader consideration, to the effect that personnel problems concerning workers' attitudes and satisfaction vary with business conditions. Procedures involving such attitudes that are effective in one period should not be applied uncritically in another.

Marital Status. Many employment men make it a practice to hire married applicants if possible, preferably those with additional dependents. The assumption is that such persons, because of their greater economic necessity, have greater incentive to do satisfactory work in order to hold the job and advance.

Practically the only available statistical studies of this factor involve salesmen. A number of such investigations are summarized in Table 52. The preponderance of married men among the superior salesmen is obvious, particularly with the higher types of selling [10, 225]. In another instance in a single company 74 per cent of the successful salesmen were married, but only 57 per cent of the unsuccessful salesmen.

TABLE 52. PERCENTAGE OF SUPERIOR SALESMEN THAT ARE MARRIED AND SINGLE⁹

Group	Married	Single
Miscellaneous.....	93	7
Insurance.....	94	6
Routine.....	61	39
House-to-house.....	81	19
Technical products.....	91	9

The results must be qualified by the fact that the majority of men of this age are married—perhaps 60 to 70 per cent. However, even if the ratio of married to single is 2 to 1, most of the ratios between married and single in Table 52 are much greater than this. Furthermore, with a group of insurance salesmen who had not been with the company over two years so that the factor of experience did not enter appreciably, the ratio of the average sales of the single group to that of the married group was \$9386 to \$10,000 [33].

Another possible source of error in the data should be noted. The married men as a rule are older. Census figures indicate that of a random selection of white men between 25 and 29 years of age, approximately 57 per cent are married, while of a similar group between 30 and 34 years about 75 per cent are married [14, 227]. We saw earlier that the best salesmen were over 30. Hence the present results may to some extent be due to the fact that the older men prove more efficient and also get married. However, three of the groups of superior salesmen listed in Table 52 show an incidence of marriage well over the census figures for other men in their early thirties. Among the routine and house-to-house groups the salaries are almost too low for the support of a family, and here are more single men. These groups are also recruited frequently from college students. But even here the married men are the larger producers.

A further refinement of analysis was made in an insurance com-

⁹ From H. G. Kenagy and C. E. Yoakum, *The Selection and Training of Salesmen*, by permission of The McGraw-Hill Book Company, Inc., New York.

pany where a number of personal history items were weighted to predict success in selling insurance and the item of marital status received a different weighting for different ages [2]. With the maturer salesmen it seemed more differential than with the young ones. So many of the latter were not married that the item seemed to have little value at that level. Incidentally, it developed that certain other factors needed to be evaluated differently for different ages. For the younger men, for example, the amount of education appeared differential, but by the time a person reached 30 other factors had apparently overshadowed the predictiveness of education. Similarly, the amount of insurance carried was more differential at one age than at another. A different selective procedure for applicants of different ages becomes complicated but may be necessary for some occupations.

As evidence from a slightly different angle we may mention the fact that among married insurance salesmen those whose wives were engaged in a gainful occupation produced only 70 per cent as much as those whose wives were dependent. The question might, of course, be raised as to which was cause and which was effect. Another insurance concern in which results like the foregoing appeared but marital condition at the time of contract was found not to be differential, discovered, however, that the greatest improvement in selling was made by the men who were single at the time of contract and had married since joining the company. This would indicate rather clearly the family incentive.

Dependents. If being married serves as an incentive for occupational effort one would expect other dependents to provide an additional motive. The same groups of superior salesmen recorded in the preceding table averaged about 2.5 dependents [10, 226]. This is more than a wife, but it is not a large family. In another company the average number of dependents of the successful salesmen was 1.9 and of the unsuccessful 0.8. Among a group of insurance salesmen those who were married but childless were slightly inferior in production to those who were single. But the production of those with 1 or 2 children, with 3 or 4 children, and with 5 or more children was in the proportion, \$10,000:\$8792:\$7584. The man with children, but with only one or two, seemed superior.

Previous Experience. It is common practice to ask an applicant regarding his previous vocational history either in general or in work similar to that proposed. Of course, if the past work has been identical with the proposed—for example, wood heeling—the case is clear. The amount of experience the applicant has had in that type of work will be somewhat indicative of his proficiency. For work at a trade it is of course desirable to develop a trade test (*infra*) instead of relying on the applicant's statement of ability or inferring proficiency from mere length of service. But even so, a statistical study may indicate that the amount of previous experience is significant. As shown above, production in selling insurance the first year was to some extent predictive of selling in subsequent years. It is not safe, however, to assume that any previous kind of selling qualifies one for a particular sales job. From the considerations in Chapter X certain types of selling apparently require a person of higher intelligence than do others. A concern found to its surprise that applicants who had had more than five years' selling experience in other lines proved to be its worst salesmen [10, 235]. It was possible that the more experienced applicants had ultimately proved unsuccessful in the other lines and then applied to this concern. Another insurance company found that its best applicants had held some other, not necessarily a selling, position for several years, but had not remained so long with a former employer as to lose their adaptability.

The kind of job previously held may give some indication of success in a proposed different line. One concern studied carefully the previous occupations of its sales force with reference to their relation to turnover, length of service, percentage of dealers sold, and percentage of quota sold. Taking all these factors into consideration, it arranged the previous occupations into seven classes on the basis of the value of the class in predicting success in selling. The order of these classes was as follows: (1) professions, (2) business for self, (3) retail selling, (4) outside selling, (5) clerical, (6) minor executive, (7) trades [10, 161]. Men recruited from the professions had a short length of service so that they constituted a rather unprofitable source of supply even though they were effective while they stayed. The next four in order constituted on the whole the best

prospective materials, but minor executives and tradesmen seemed a rather unprofitable source from which to recruit for this particular selling job.

Some systematic studies of this problem of experience have been made by the U.S. Employment Service with a view to discovering groups of jobs requiring similar physical and mental characteristics. For instance, we may take a considerable number of jobs that are clerical in nature and think of them as one general family. We may then take samples of people from several of the jobs in this family, select psychological tests, and validate them on the entire sample. In this way it may be possible to develop a test for clerical workers in general. While it undoubtedly would be better from a scientific standpoint to have separate tests for punch-card operators, coding clerks, private secretaries, bookkeeping-machine operators, transcribers, and calculator operators, it is of some value to have a battery of tests for the clerical occupations in general. Another instance is tests developed for department store salespeople. Obviously, they are selling many lines of goods where the problems may be considerably different, but it is feasible to draw samples from a wide range of sales departments and validate the test on this heterogeneous sample of salespeople.

Another expedient is more in line with the present discussion where tests are not under consideration. It may be possible to *estimate* the characteristics that are necessary for the job. This is done in connection with the job analysis. If, for example, the job requires a lot of lifting, obviously strength in the back is needed. Similarly, it might be fairly evident to the analyst that the worker had to distribute his attention to a considerable number of things, that he had to remember numbers or symbols, or to judge distances. If a procedure of this sort is carried through for a number of occupations it may be possible to determine those that have certain characteristics in common. Then if an individual has had some successful experience in one kind of work there is a presumption that it is safe to employ him in another job which has similar mental aspects. The methods of making a job analysis are discussed in detail in a later chapter. A check list may be helpful, including items like dexterity of fingers, estimation of speed, memory for directions, emo-

tional stability. The analyst may indicate the amount of each characteristic demanded of the worker in arbitrary units such as A, B, C. As the result of these analyses it may be possible to group jobs together on the basis of some of these necessary characteristics. For example, a group of 26 jobs were found that required a B grade of dexterity but no minimum formal education or special knowledge or experience, and that were repetitive. These occupations included bagger, carton closer, capper, sticker, mold filler, nailer, fur glazer [31, 191].

Further light is sometimes thrown upon these occupational patterns by studying occupational histories. If, for instance, many workers are found in Job A who have had previous experience in Job B, it suggests that these two jobs have something in common. As a result of such analysis it may be possible to make up occupational patterns as above described.

Miscellaneous Factors. Certain miscellaneous items of personal history may be of significance in a particular situation. For instance, with insurance salesmen the number of clubs to which a man belonged was somewhat indicative of production. The correlation coefficient is small, but the amount of paid business solicited increases gradually with increasing number of clubs; the men belonging to seven clubs had the best record of all [17]. With insurance salesmen those who carried a considerable amount of insurance themselves proved more effective. Reasons for entering a vocation may have some significance. It was found that employees who entered an occupation because of the influence of a friend were not as effective as those who had entered for ulterior reasons. Possibly this latter reflected a real interest in the work, whereas the former indicated mere accident.

Combinations of Personal History Factors. In some cases efforts have been made to determine roughly the validity of various items in the personnel blank and combine them into a weighted score [5]. For an insurance company certain items were weighted as indicated in Table 53. The weighting takes account of the fact that very young persons are not as apt to be successful in selling as are the middle-aged. Similarly, with education there appears to be an optimum value at about twelve years of schooling. Married applicants receive more consideration than unmarried. Previous occupation seems significant when considered

TABLE 53. WEIGHTS OF ITEMS OF PERSONAL HISTORY FOR PREDICTING EFFICIENCY IN SALESMANSHIP¹⁰

	Weight		Weight
Age:		Education:	
18 to 20	-2	8 years	+1
21 to 22	-1	10 years	+2
23 to 24	0	12 years	+3
25 to 27	+1	16 years	+2
28 to 29	+2	Occupation:	
30 to 40	+3	Social	+1
41 to 50	+1	Non-social	-1
51 to 60	+0	Own insurance:	
Over 60	-1	Carried	+1
Marital status:		Not carried	-1
Married	+1	Contract:	
Single	-1	Full time	+2
Clubs:		Part time	-2
Belongs	+1		
Not belong	-1		
Experience:			
Previous life insurance experience			+1
Confidence:			
Replies to question: "What amount of insurance are you confident of placing each month?"			+1
Does not reply to this question			-1

from the standpoint of whether or not the occupation involved social contacts, such as selling, work at a cashier's window, or reporting. It also appears that individuals who are contemplating full-time service are a better investment than those who propose to work only part time. Carrying insurance oneself is likewise apparently in the applicant's favor as is also belonging to various clubs.

When a large group of insurance salesmen were classified, on the basis of their production records, into a best group, a middle group, and a poorest group, and their scores on the personal history blank were computed according to the above weighting, the results were as shown in Table 54. The entries in the table are the percentage of individuals in the given production group falling in the various classes of weighted score. The largest percentage of the best group scores above 8 points;

¹⁰ After Goldsmith.

TABLE 54. PERCENTAGE OF SALESMEN WITH DIFFERENT PRODUCTION RECORDS MAKING VARIOUS SCORES ON WEIGHTED PERSONAL HISTORY ITEMS¹¹

Production Record	Score on Personal History Blank		
	Below 4	4 to 8	Above 8
Best group.....	15	41	44
Middle group.....	18	54	28
Poorest group.....	53	37	10

much smaller proportions of the two other groups do so. The poorest group has the majority of its scores below 4. On the basis of these results a critical score of 4 points was recommended. If applicants below this score were rejected, it may be seen that many of the inferior salesmen would be avoided and comparatively few of the better ones eliminated.

Another life insurance company finds that the following ten items have differential value in selecting their salesmen: marital status, education, previous income, life insurance owned, previous occupations, selling experience, minimum living expenses, length of residence in the community, present membership in organizations, length of time of negotiations prior to employment [24].

An investigation was made of 172 YMCA secretaries. They were ranked by five judges, extreme groups were selected, and biographical items were investigated as to the percentage of these groups that manifested certain characteristics. The most differential items were selected and weighted. These included such items as farm or city birthplace, academic record, special studies, number of children. A critical score was derived such that if it had been applied to 18 men who had left this vocation, the fact of leaving could have been predicted for 15 of the 18 [1].

A cooperative organization sponsored by a number of insurance companies has been at work for some years developing techniques for the selection of agents and reporting privately

¹¹ After Goldsmith.

to the companies participating. As a criterion they have used production and the fact that an individual has stayed with the firm for two years. One of their documents furnishes a list of personal history items with appropriate weights set up in comparatively fool-proof fashion for the use of any agency. Details cannot be given because of the confidential nature of the material, except to indicate how the technique resembles that in some of the other studies reported. For example, no dependents receives a weight of 3; one dependent, 4; 2 dependents, 6; 3 dependents, 8; weights decrease beyond 4 dependents. Membership in no organizations or only one is given a weight of 3; membership in 2 increases the weight to 4, and membership in 3 makes it 8. Other items include previous occupations, offices held in organizations, length of time with present employer, amount of insurance carried. All the data are available from the personnel blank. After the total has been obtained, an age adjustment must be made for certain characteristics because some of the items show a positive correlation with age. For example, as a man grows older he has more dependents; hence, to be equally indicative of success, the number of dependents must be a little larger for a middle-aged person.

It is sometimes possible to apply correlation procedure to items of personal history and weight them according to a regression equation just as was done with tests of special capacity. This is, of course, feasible only when the factors involved are such as yield a considerable range of possible values. A correlation based on a variable that involves only two classes, such as married vs. single, is not well adapted to this procedure. Such an equation for insurance salesmen proved to be:

$$X_1 = 3.2X_2 + 9317X_3 + 106X_4 + 5534X_5 + 26880$$

where X_1 is production, X_2 the amount of insurance carried at the time of contract, X_3 the number of clubs to which the man belongs, X_4 the age at the time of contract, and X_5 the number of dependents at the time of contract [17]. When the items are weighted according to this equation the coefficient of multiple correlation—i.e., the correlation of the weighted sum of these items with the criterion—proves to be .40. This is a considerably better prediction than could be made with any single item.

A correlation like this is not, of course, sufficiently high to justify its use as the sole basis for selection or even in lieu of the various tests that might have been developed. However, such a weighted personal history record might form a valuable supplement to any other predictive measures that were available. If the data are in a form so that correlation coefficients can be computed and a regression equation worked out, the effort may prove worth while. In some instances it may be possible to include some of these personal history items in a regression equation along with tests.

The foregoing are some of the items of personal history that are available in the average application blank and that have been shown in some situations to be indicative of occupational success. As with many other predictive measures, one cannot assume that what has proved valid in one situation will do likewise in a different one. It is necessary to evaluate the items with reference to the special situation in which they are to be used.

A brief description may be given of a personnel blank at the academic level. This blank is used by a large psychology department in the process of selecting junior staff members. The first item deals with general scholarship, giving in addition to institutions and degrees a breakdown of grades obtained in each year of college or graduate school. The data are analyzed separately from the standpoint of all courses and psychology courses. The next item is scholarship in the department of psychology; the courses taken are subdivided into general, experimental, comparative, educational, clinical, etc., with the course titles and grades. The third item is intelligence. The fourth is background courses of particular use for psychology under such headings as physiology, statistics, mathematics, curriculum construction, guidance, physics, chemistry. The fifth item deals with research and calls for information as to field of specialization, topics for M.A. or Ph.D. thesis, and any researches completed, in progress, or published. Item 6 concerns special abilities and includes a list of about 60 abilities—such as alphabetizing, administering group tests, compiling bibliography—which the subject checks as to whether he is expert, skilled, familiar, or

novice. The seventh item deals with his attitude, purpose, ambition, and study program—with information as to his plans after completing his education, his vocational interests to date, and the probable schedule according to which he will finish his work. Item 8 deals with teaching experience, including the place, time, and subject, and the name of the immediate superior; the latter reports on a separate blank. The ninth item concerns organizing and executive ability and includes questions as to the most important responsibilities held, societies organized or managed. Item 10 is aimed at social capacity and calls for information regarding extracurricular activities in high school and college, offices held in organizations, present membership in societies, selling experience, public speeches. The eleventh item deals with mechanical ability and experience, such as apparatus designed or constructed or machine shop experience. Item 12 deals with experience as an assistant in academic institutions. Blanks for miscellaneous personal data follow regarding age, sex, marital status, dependents, health. On the concluding page the applicant writes a sketch of his life.

These items were selected originally by a committee that followed essentially the procedure discussed earlier for selecting and weighting items for rating scales. The members of the committee listed the qualifications which they thought necessary and secured the judgment of other people regarding their importance. After the list had been reduced to its final number each committee member spent 100 points among these items; the approximate average weights were obtained in this fashion. After the blank had been in use for some years revisions were made by comparing item scores with what was known about the subsequent success of the individuals who had been employed on the basis of this blank and also by determining which items had the highest correlations with total score on the blank [37]. In fact, the blank was subsequently rearranged with the items in roughly the order of their validity so that if an applicant showed marked inferiority in the first few it would not be necessary to carry through detailed evaluation of all the remaining ones. The raw weights assigned to the different items are as follows:

Research	10
Years beyond A.B.	2
Grades in psychology	10
Grades in general	7
Number of hours psychology	4
Number of areas	3
Intelligence	10
Background	7
Special abilities—psychology	6
Special abilities—other	2
Languages	3
Years teaching	5
Quality of teaching	10
Years as assistant	5
Social	10
Organizing and executive	6
Attitude and purpose	4
Mechanical	4
Disabilities (absence)	5

When the blanks were actually used, some of the items yielded objective scores and some involved ratings on a 10-point scale by the committee. By the use of standard score procedures, items could be equated and then weighted according to the raw weights given above.

LETTERS OF APPLICATION

The first step in many employment situations is the solicitation of letters of application. Help-wanted advertisements often require this form of reply. Such letters serve a purpose similar to that of the application blank in enabling a preliminary sorting of applicants with a view to finding those in whose case interview or further investigation is desirable. If a grossly misspelled letter is received from an applicant for a stenographic position the matter ends right there. It is also possible in this way to get a line on individuals who are at a considerable distance and do not care to come and apply personally unless there is a fair prospect of their being hired.

A letter of application differs from the usual application blank in that it insures less in the way of specific information. Instead of the applicant being asked specifically for biographical data

he simply writes the facts which he considers most pertinent in qualifying him for the position in question. This factor of giving the applicant a chance to express himself freely, although he may omit some significant items, makes it possible in the opinion of some employment men to judge something regarding such matters as neatness, ability to express oneself, or tendency to be systematic that might not be manifest in the answers to specific questions. However, this feature can be included in a standard application blank—the one for academic positions just described included as the last item a life sketch.

In considering letters of application it is necessary to make some qualitative evaluation of them. The average employment man deals with them by the usual procedure of general impression. One letter may be markedly neater than another, thus leading to the presumption that the writer of the first is the neater individual. However, a question arises as to the reliability and validity of such estimates of the individual from the letter. The reliability of the estimates involves the extent to which a given judge agrees with himself if he makes the estimate on different occasions, or the extent to which he agrees with other judges. The validity of the estimates denotes the extent to which the results correlate with a further criterion, such as production or the judgment of persons who are acquainted with the applicant and are not judging him merely by his letter. These problems of reliability and validity of estimates suggest the earlier discussion of evaluation of estimates of mental traits from physiognomy as manifested in photographs. (Cf. Chapter II.)

The Reliability of Estimates. Experiments on the reliability of estimates have been conducted. An advertisement for a book-keeper and office assistant was inserted in a New York paper and 25 of the letters of application were selected for study [7, 10]. The signatures were removed and the letters marked with a key symbol. These letters were then submitted to 50 judges—business and professional men and women, students, and clerical workers. These judges ranked the letters in order from best to worst; i.e., they numbered them from 1 to 25, with reference to (1) intelligence, (2) reliability, (3) tact, and (4) neatness. The ratings on these four traits were made separately. In addition,

10 of the judges repeated this same procedure a month later without referring to their earlier estimates.

A detailed presentation of the results is not worth while in the present connection. Suffice it that there was a rather marked disagreement among the judges. For simplicity's sake we shall consider only 10 typical judges. The first letter, when estimated as to the intelligence of the writer, was ranked all the way from 3 to 17. This same letter received a rating of from 2 to 23 for tact; from 4 to 19 for reliability; and from 1 to 13 for neatness. Considering the next letter in the same fashion, the ranks assigned by these 10 judges varied for intelligence from 4 to 45, for tact from 1 to 24, for reliability from 4 to 20, and for neatness from 4 to 25. These results were by no means atypical of those with the other 40 judges. If they had shut their eyes while considering the tact or reliability of the writers, they would have agreed with one another regarding the order of the letters about as well as they did with their eyes open. The situation is slightly improved with reference to neatness and intelligence, but not to any great extent. Thus the reliability of the estimates from the standpoint of the agreement of the judges with one another seems to be rather low.

To study reliability from the other standpoint of agreement of the judge with himself, we may consider the results for the 10 judges who repeated the ranking procedure a second time one month subsequently and correlate their two sets of estimates. Such correlations will show, for instance, whether a given judge ranks the same letter high in intelligence in both first and second trials and rates another letter low in intelligence in both cases. These correlation coefficients are given in Table 55. This table shows, for example, that Judge A's initial ranking of the letters from the standpoint of intelligence correlates .59 with his ranking a month later. However, his initial and final rankings for tact correlate to the extent of only .40. The average of his four correlations is .54. This average gives a fair notion of the reliability or consistency of Judge A. A glance at the figures in the table shows that Judges B, D, and I are rather effective from this standpoint, whereas C and F are distinctly inferior. If they were hiring employees on the basis of letters of application, many a person's destiny would hinge on the weather the

TABLE 55. CORRELATION BETWEEN ESTIMATES OF TRAITS FROM LETTERS OF APPLICATION MADE ONE MONTH APART¹²

Judge	Intelligence	Tact	Reliability	Neatness	Average
A.....	.59	.40	.50	.67	.54
B.....	.72	.72	.73	.72	.72
C.....	.08	.40	.27	.38	.28
D.....	.72	.44	.65	.88	.67
E.....	.60	.63	.20	.44	.47
F.....	.31	.18	.23	-.14	.21
G.....	.44	.52	.46	.92	.60
H.....	.62	.31	.45	.51	.47
I.....	.65	.71	.73	.91	.75
J.....	.63	.42	.52	.71	.57
Average.....	.54	.47	.47	.60	.53

day his letter happened to arrive or on what the personnel man had eaten for lunch.

The results may also be considered from the standpoint of the agreement of the individual judges with the consensus of all the judges. Some apparently agree more closely with the consensus than do others, but it proves rather difficult to locate an expert, i.e., one whose individual opinion is tantamount to the combined opinion of all. The results of this study, then, are discouraging from the standpoint of the consistency of the judges' estimates based on letters of application.

It is probable that the unsatisfactory character of the results is partly due to the halo effect mentioned in the preceding chapter. As a matter of fact, in the present study it was found that there was a high intercorrelation between the traits. The correlations of intelligence with tact and reliability and of tact with reliability are over .90, and all the other correlations are over .80. Evidently the ratings were largely a matter of general impression.

The Validity of Estimates. One experiment on the validity of estimates based on letters of application may be described [26].

¹² From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, New York.

Twenty-five seniors in a school for religious workers wrote personal letters of application for positions of the type for which they were preparing. These letters were submitted to 12 members of the faculty at the Union Theological Seminary who ranked them according to the degree to which the individual's letter indicated "general fitness for the position." To obtain a criterion by which to evaluate these estimates, 5 of the students' teachers ranked them for general ability, intelligence, and tact. They were similarly rated by one another—each student ranking the other 24 members of the group in these traits. It is interesting to note that the teachers' and the student associates' ratings correspond closely for general ability with a correlation of .90, and for intelligence with a correlation of .83. The correlation for tact is .59.

The real problem, however, is the extent to which the estimates made from the letters of application agree with those made by teachers or associates who were acquainted with the individuals. These correlations are given in Table 56. For instance, the correlation between general fitness for this type of

TABLE 56. CORRELATION BETWEEN ESTIMATES OF GENERAL FITNESS FOR A POSITION BASED ON LETTERS OF APPLICATION AND ESTIMATES OF TRAITS BY ACQUAINTANCES¹³

	Teachers	Associates	Teachers and Associates
General ability.....	.56	.46	.50
Intelligence.....	.58	.44	.44
Tact.....	.20	.18	.22

work in the opinion of those evaluating the letters and general ability as estimated by the teachers who were in contact with these applicants is .56, whereas the correlation between this estimate from the letters and the judgment of student associates as to general ability is .46. When the judgments of teachers and associates for each individual are combined into a single judgment of general ability, these figures correlate with the estimates

¹³ After Poffenberger and Vartanian.

from the letters to the extent of .50. Similar correlations are given for the other traits.

It is to be noted that there is a fair correlation between estimates of general fitness based on the letters and estimates by acquaintances as to general ability and as to intelligence. These correlations, however, are lower than those obtained in many projects using tests for vocational prediction. The estimates of tact are apparently of no value as far as correlation with the criterion is concerned.

It is interesting to compare the results obtained by pooling the estimates of all the judges with those obtained by the individual judges. If the results for each of the 12 judges are correlated separately with the criterion, the three traits being combined into a single figure for a given judge, the average of these correlations is .37. If, on the other hand, the estimates for all the judges are combined in a single figure for each candidate and these pooled estimates are correlated with the criterion, the coefficient is .46. This is the same tendency that has been found in other connections, namely, that better results are obtained by combining the estimates of a number of judges than by using the estimates of any particular judge. The correlation between the criterion and average estimates is usually larger than the average of the correlations between the criterion and individual estimates. This suggests the possibility, if members of a staff are evaluating application letters, of adopting a technique whereby they independently rate the letters and then combining these ratings into a single figure for each letter.

Graphology. A word should be said regarding graphology in the present connection, because some employers may have the idea that they can infer various character traits from the handwriting in the application letter. Most of the generalizations in this field are based on analogy. It is assumed that writing continuously from letter to letter denotes coherent thought, while breaks between the letters indicates that the person is addicted to flashes of inspiration; that heavy writing denotes strength of will and persistence; that large bold writing denotes a person with imagination and ambition. These conclusions are not based on empirical evidence. Even such reasoning as that neatness in writing connotes general neatness is unwarranted. Habits are

specific rather than general. Ambition to win in golf does not necessarily denote desire to do one's best in the factory. Enthusiasm for social contact at a dance differs from desire to meet people from behind a cashier's window. Neatness in handwriting or in personal appearance is not a universal index of neatness in clerical work.

When the alleged assumptions of graphology are evaluated statistically the results are similar to those found by similar statistical studies of the claims of physiognomy. (Cf. Chapter II.) One such study included the following claims of graphology: Ambition is associated with upward-sloping lines, bashfulness with fineness of lines and narrowness of *M*'s and *N*'s; forcefulness is associated with heavy lines and heavy bars on the *T*'s, perseverance with long bars on the *T*'s, and reserve with closed *A*'s and *O*'s [9]. Seventy students in a medical fraternity rated one another on these traits. Each one copied a piece of prose on the same kind of paper with the same pen and these handwriting specimens were measured in detail, sometimes with the aid of a microscope. The correlations between handwriting data and the estimates made by acquaintances ranged from .38 to -.20, and two-thirds of them were negative. In another study, judges estimated the intelligence of a large number of students on the basis of samples of their handwriting in a uniform piece of dictation. The estimates were compared with results in an intelligence test. The correlations for the different judges ranged from .16 to -.16, indicating utter inability to judge intelligence from handwriting [21].

There are suggestions that sex can be judged correctly from handwriting a little more frequently than chance expectation. A review of a number of such studies [38] indicates that when miscellaneous samples of handwriting are judged as to the writer's sex the judgment is correct in a little over 60 per cent of the cases. However, the personnel man does not need a technique for judging the sex of the writer of an application letter.

Estimates of Oneself. Inasmuch as the writer of a letter of application often gives some evaluation of his own traits or capacities, it is in point to consider how well one *can* evaluate himself. Studies in which persons have rated themselves in various traits and their ratings have been compared with the

ratings of intimate acquaintances reveal the tendencies [7, 49]. In one instance 25 people ranked one another, themselves included, in a list of traits. For a given trait the average rank assigned an individual by his associates was taken as his actual possession of the trait. It was then possible to note how the rank he assigned himself deviated from this "true" rank. These results are shown in the first column of Table 57, which gives the average of such figures for all the subjects. For purposes

TABLE 57. ESTIMATES OF ONESELF COMPARED WITH ESTIMATES BY ONE'S ASSOCIATES¹⁴

	Average Displacement of Self-estimates from Estimates of Associates	Average Deviations of Estimates of Associates	Average Overestimation of Self
Neatness.....	5.8	4.5	+1.8
Intelligence.....	6.0	3.7	+3.0
Humor.....	7.3	4.5	+5.2
Conceit.....	5.7	4.1	-1.7
Beauty.....	6.0	3.8	+0.2
Vulgarity.....	6.1	3.5	-4.2
Snobbishness.....	5.1	4.8	-2.0
Refinement.....	7.2	5.9	+6.3
Sociability.....	5.4	4.7	+2.2
Average.....	6.1	4.4	

of comparison there was also computed the average agreement among the judges in estimating each trait. These figures appear in the second column. In this case a purely chance arrangement would give an average deviation of a little over six steps. The self-estimates in general deviate to almost this extent, while the estimates by acquaintances are appreciably better. The results for another group of individuals who performed a similar experiment are given in the last column of the table. It shows the tendency to overestimate (+) or underestimate (-) one's traits relative to the average estimate of acquaintances. The tend-

¹⁴ From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, New York.

ency seems to be to overestimate oneself in the more desirable traits and to underestimate oneself—i.e., give oneself a higher rating than one deserves—in the undesirable traits. Consequently, statements regarding an applicant's mental traits in his own letter, even though sincere, are of dubious value.

RECOMMENDATIONS

Difficulties. Recommendations are required in many employment situations. When a prospective employer does not know the applicant personally, it seems perfectly logical to make inquiry of someone who does. If the former employer to whom inquiry is made is serious and fair and his ability to judge traits reliable, his recommendation should be of some actual value, but there are difficulties with the procedure on these very points. The first of these, about which unfortunately it is impossible to obtain scientific evidence, is the bias or carelessness of the writer of the recommendation. The recommendation is often "sealed with a shrug and opened with a smile." The recommender may be led to overstate the case through his desire to facilitate the exit of a present employee. On the other hand, he may wish to keep the employee and to this end may understate the case. In still other instances the recommender may have no particular bias but may make inadequate statements of a perfunctory character. Many recommendations are of this sort. The recommender feels some doubt as to the value of the whole procedure and as to his ability to evaluate the candidate and he uses certain set or conventional terms on all such occasions. In such instances the apparently detrimental content of the recommendation reflects not the applicant's lack of ability, but rather the recommender's apathy regarding the applicant's destiny.

Suppose, however, that the writer of the recommendation is unbiased and tries to do his best in evaluating the applicant; other possible sources of error should be considered before attaching much significance to his statements. Something depends on the aspects in which he is called upon to evaluate the applicant. The preceding chapter showed the necessity of carefully working out the details of a rating procedure and of training the raters if any great value is to be achieved in considering character traits. This has obvious implications regarding the

value of estimates made by untrained persons in writing a recommendation. Some aspects, however, may not be as bad as others. It has been pointed out that the more objective traits are rated somewhat more reliably than the subjective. In evaluating a recommendation, perhaps greater significance should thus be attached to statements regarding objective traits.

Another thing that should be considered is the relation between the one making the recommendation and the applicant, with special reference to the conditions under which the former has generally observed the latter. If the applicant has been a pupil or parishioner of the recommender, the conditions of observation will be quite different from what they would have been if he were an employee.

The conditions under which a trait is judged makes some difference in the reliability of its estimation, as the following study shows [7]. A group of teachers rated one another in seven different traits; a group of students rated one another in the same traits; and finally a group of students judged their teachers in these traits. In each instance the reliability of each trait was determined by computing the agreement of the judges with each other. The results appear in Table 58. The actual deviations are not shown, but merely the relative order of reliability for the traits. For example, with teachers judging teachers efficiency is rated with the most reliability, while with students judging students the most reliable estimates are for independence. There is obviously a fair correspondence between the relative reliability of the traits when teachers judge teachers and when students judge students. The results are quite different when the students judge the teachers. Some of these reversals are understandable. Estimates of kindness and cheerfulness, for instance, are most reliable for the students judging the teachers and much less reliable for the teachers judging one another. Kindness is a trait that the students would collectively have a chance to observe in the classroom, and the same thing would be true of cheerfulness. Under these circumstances the students would therefore make rather uniform judgments of these traits inasmuch as they would observe them operating under the same conditions. The teachers judging one another, however, do not make their judgments under uniform conditions, and one of them will see a man

TABLE 58. ORDER OF RELIABILITY OF ESTIMATES OF TRAITS¹⁵

	Teachers Judging Teachers	Students Judging Students	Students Judging Teachers
Efficiency.....	1	2	5
Energy.....	2	4	3
Leadership.....	3	3	7
Independence.....	4	1	6
Cooperativeness.....	5	6	4
Cheerfulness.....	7	5	2
Kindliness.....	6	7	1

in a situation where his kindness will manifest itself and another in a situation where there is no such opportunity. On the other hand, leadership is rather poorly judged by the students. It apparently does not manifest itself in the classroom situation. Fellow teachers, however, have rather common criteria in the social environment by which to estimate the leadership of the teacher in question and do so more effectively than students. The point then is that estimates of traits depend for their value to quite an extent upon the relation between the judge and the judged. It would seem offhand that the recommendation of a former employer who had observed the individual in the actual industrial situation would be more valuable than that of a person whose observation had been confined to other situations.

Kinds of Recommendations. There are three general kinds of recommendations. The first is the testimonial which the applicant solicits and takes with him when leaving an employer. This type of recommendation is usually a brief statement of satisfactory service. It cannot go into much detail or give anything of a confidential character because the applicant sees the letter himself. It could not say: "To whom it may concern: The bearer, Mr. John Doe, is a crook," even though the statement is warranted. A second type of recommendation consists of a letter written directly to the prospective employer at the applicant's request. This is better than the first type because it involves

¹⁵ From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, New York.

confidential material; the previous employer can write without restraint and, if he wishes, can give an unbiased account of the individual's qualifications. The third type of recommendation consists of a letter in response to an inquiry from a prospective employer to a previous employer. This has the great advantage of calling for, and probably obtaining, the specific information that is wanted. Whereas in the other cases the prospective employer may receive a lot of high-sounding irrelevant material, in this case he obtains information primarily on the points which he considers significant in his particular situation.

The last of these types of recommendation is the only one that is worth serious scientific consideration. The conventional method is to write a simple personal letter of inquiry, but there is the possibility of some refinements in this method. It is feasible, for instance, to construct the inquiry in such a way as to save time for the one answering it as well as for the one who will subsequently evaluate it. While specific questions that require sentences for an answer may be asked, the same information can be obtained by having the reader indicate his answers by a few check marks or at the most by a few words. The following blank is typical:

Dear Sir:

Mr. has applied to us for a position as
 and has named you as a former employer. It will help us
 if, in entire confidence, you will give us the information requested
 below. We shall be glad to reciprocate at any time.

1. In your opinion is he honest and responsible? Yes.... No....
2. Is he temperate with tobacco and alcohol? Yes.... No....
3. Does he possess skill in the work named above?
 High skill..... Generally qualified..... Doubtful.....
 No skill.....
4. He states that he was in your employ as.....
 from.....to..... Does this correspond to your
 record? Yes.... No....
5. He states that he left because
 Is this an adequate statement? Yes.... No....
6. He states that he received in salary or commission.....
 per Is this correct? Yes.... No....
7. Would you reemploy him? Yes.... No....

8. If not, will you please give reasons
9. If you have further information that will assist us in helping him make the most of his opportunity, kindly indicate it
10. If you have further information that can better be given in personal communication with our representative, check here.....

For obtaining information dealing more specifically with various traits a scheme somewhat similar to a rating scale has sometimes been used. After an introductory statement like the preceding, the reader is requested to check or ring the word in each line that most nearly describes the applicant.

<i>Physical appearance</i>	Commanding Insignificant	Pleasing	Average	Unattractive	
<i>Clothes</i>	Stylish	Well dressed	Ordinary	Untidy	Shabby
<i>Manners</i>	Obtrusive Bashful	Friendly	Well mannered	Retiring	
<i>Ambition</i>	Keenly ambitious Easily satisfied	Moderately ambitious Indifferent	Lacking		
<i>Application</i>	Exceptionally industrious work assigned	Shiftless	Industrious Lazy	Performs	
<i>Persistence</i>	Very persistent discouraged	Determined A quitter	Ordinary	Easily	
<i>Popularity</i>	Very popular Unpopular	Good mixer	Average	Exclusive	
<i>Parents</i>	Wealthy Working people	Well off Poor	Moderate circumstances		

The above items are, of course, merely suggestive and would necessarily vary with different occupations. However, recasting the recommendation blank into this form enables the prospective employer to obtain the desired information with a minimum outlay of time on the part of the one filling out the blank and the one evaluating it. If an individual is repeatedly solicited for recommendations which he must answer at length, he naturally drops into perfunctory habits. If, however, the request is presented in such a fashion that he can check the answers in a very

few minutes, he will react to it much more favorably and be more apt to exercise his best judgment.

A recommendation blank used in the selection of graduate assistants in a large psychology department may be described. It supplements the personnel blank described above (p. 434). It is not in graphic rating scale form but it requires a short statement about the individual on a number of points. The kind of information desired is indicated by a few questions. Some of these items follow.

Research Promise. (The person's application blank furnishes a list of researches and publications.) What is your evaluation of his past and present researches and publications and his future research promise? What will be his research status ten years hence? Will he be likely to hibernate? What evidence has he given of originality, and grasp of scientific method?

Special Abilities (useful for this position): e.g., unusual statistical or mathematical ability, related sciences (neurology, anatomy, genetics, chemistry, physics, etc.), shopwork, engineering, laboratory or clinical experience, teaching, travel, drafting, typing, stenography, photography, artistic ability, calculating-machine operating, mimeographing, grading papers, scientific background, etc., according to the position applied for.

Attitude, Purpose, Ambition, Study Program. Does applicant have a Purpose and a Program? Is his present interest in the major department, mentioned above, probably a permanent one; or would it be adversely affected by failure to receive appointment? Has he given up any career for which he was obviously unfitted, or which was incompatible with his present plans and ambitions? How long has he had his present ambition?

Teaching Experience and Ability. (If no experience, what is your estimate of the applicant's promise as a teacher?) If employed as a teacher, has he been offered, or will he be offered, reappointment? If your funds permitted would you be willing to employ or to reemploy him as a teacher?

Social Capacity. Has applicant the ability to direct others and to gain the cooperation of students and associates without antagonizing them; to "sell himself" to a prospective employer; to make a favorable impression before a class? Is he physically attractive; optimistic; fundamentally interested in people; competent in the social amenities?

Other items on the blank are scholarship, intelligence, organizing ability, mechanical ability, special disabilities, weaknesses that need

improvement. After each item there is a blank space of about an inch in which the recommender writes his statement.

THE INTERVIEW

Employees are seldom hired without a personal interview by some member of the staff. In the first place, there is a rather general feeling that it is desirable to see the applicant and talk to him with a view to sizing up certain traits that might not be revealed by any other procedure. In the second place, the interview may give the applicant information about the nature of the proposed work and about the company so that his subsequent experience will not run counter to his initial impressions. In the third place, the interview affords an opportunity to make a friend for the company so that the applicant will desire to work for it.

The first of these functions is the one that has received the greatest stress and experimental study. If the information as to the applicant's mental or other qualifications revealed by the the interview is valid, this constitutes, of course, a convenient and expeditious method of hiring. Many executives, however, have a probably unfounded confidence in their ability to predict occupational success by this method. At any rate, the interview is such a common practice that it is desirable to investigate its worth scientifically, particularly from the standpoint of the value of judgments regarding the applicant's qualifications. Interviews vary widely in character. In some cases a rather perfunctory set of questions is asked with a view merely to keeping the applicant engaged so that he can be watched. In other cases the questioning is more flexible and exhaustive, with a view to obtaining as much information as possible about the applicant's qualifications.

Factors Making for Unreliability. The customary interview procedure is not all that can be desired, for there are a number of psychological factors that tend to produce unreliability. In the first place, the interviewer is prone to use personal generalizations about such things as physiognomy. (Cf. Chapter II.) If, for instance, one has had an unpleasant experience with some person who has a long nose or red hair, he is likely to impute the same unpleasant characteristics to an applicant with these physiognomic aspects. Many people have some almost uncon-

scious generalizations of this sort which doubtless considerably influence their judgment of people whom they observe. Allusion was made earlier to the school principal who selected boys largely on the basis of the way they walked down the aisle toward his desk.

To be sure, such generalizations may occasionally be sound and based on psychological principles, but the difficulty is that it is impossible to ascertain whether or not they are sound unless recourse is had to statistical methods. And even though they are sound, one interviewer may be using one method and another interviewer a quite different one. This would account for discrepancies between interviewers, and for the fact that occasionally an interviewer shows unusual facility at his task although he may not know exactly how he does it. In fact, some of the commercial character analysts who purported to be using physiognomy and were doing a good job of it evidently were actually using their own background of experience without realizing exactly what they were doing, for when someone else attempted to use the alleged physiognomic system it did not work. At any rate, it is well for the interviewer to note whether he is using any generalizations that are based largely on his personal experience and have no scientific validity, unless he knows that they have some statistical foundation.

A second factor making for unreliability of the interview is the frequent assumption that habits are general rather than specific. It is assumed that a habit formed in one field with reference to one kind of situation will operate in other fields—for example, that an applicant who is neat in dress will likewise be neat in work, or that one who talks rapidly and seems very much alive will be a rapid worker, or that one with awkward physical posture will be inaccurate and clumsy in manual work. As a matter of fact, habits are not usually generalized to this extent; they are more frequently specific in character, and are correlated with specific pathways in the nervous system. The habit, for instance, of looking in a mirror and adjusting the necktie deals specifically with the motions involved in adjusting the tie and not with the motions involved in making a micrometer adjustment on machine tools. The neural pathways in the two instances are quite different. Again, the neural pathways that lead to the speech muscles

of a rapid talker do not lead to his hands and give a presumption that he will be rapid in manual work. Similarly, a person who is clumsy in the control of his larger muscles, feet, and arms may not be equally clumsy in the fine coordinations of his fingers in doing delicate machine work. It is easy to conceive special cases in which, even if the habit were somewhat generalized, other factors would enter to break it down. An applicant for an executive position might present himself with grimy hands. This might reflect not personal untidiness, but rather the fact that while waiting for a position in his own line he took a mechanical job as the next best thing. The particular traits leading him to make such a shift might really be the kind that would make him better qualified for the executive position. In other instances a person might be somewhat unkempt because of such tremendous interest in an invention or a piece of research that he was conducting that he temporarily neglected his appearance.

A third factor that contributes to the unreliability of the interview is the "nervousness" of the applicant. It is quite possible that many an applicant, in the excitement of the situation, particularly if it is very important for him, will be in an abnormal mental condition. An individual who is usually fairly calm may under these circumstances show what seems like distinct nervous instability. In giving tests it will be recalled that a "shock-absorber" test often precedes the tests proper to alleviate this initial emotional disturbance. A skillful interviewer will probably be able in the course of the conversation to determine whether or not the applicant is in such a state; if so, he should be able to remedy the condition. There are times, of course, when nervousness during an interview reflects a fundamental characteristic of the applicant's personality and not a temporary condition. A good interviewer should be able to detect the fact.

Demonstration of Unreliability. While the foregoing factors presumably make for unreliability of the interview, it is further possible to study the matter statistically just as reliability has been studied in other connections. Fifty-seven applicants for sales positions were interviewed individually by 12 sales managers [7, 65]. These managers were allowed to conduct the interviews in whatever fashion they wished, but at the conclusion they were required to rate the individual as to "suitability for

the position in question." These ratings were then cast into such form that each applicant could be assigned a rank from 1 to 57 for each judge. The results for a typical group of applicants are

TABLE 59. RANKS ASSIGNED APPLICANTS BY SALES MANAGERS WHO INTERVIEWED THEM¹⁶

Applicant	Sales Managers												
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Range
A.....	33	46	6	56	26	32	12	38	23	22	22	9	6-56
B.....	36	50	43	17	51	47	38	20	38	55	39	9	9-55
C.....	53	10	6	21	16	9	20	2	57	28	1	26	1-57
D.....	44	25	13	48	7	8	43	11	17	12	20	9	7-48
E.....	54	41	33	19	28	48	8	10	56	8	19	26	8-56
F.....	18	13	13	8	11	15	15	21	32	18	25	9	8-32
G.....	33	2	13	16	28	46	46	32	55	4	16	9	2-55
H.....	13	40	6	24	51	49	49	52	54	29	21	53	6-54
I.....	2	36	6	23	11	7	7	17	6	5	6	9	2-36
J.....	43	11	13	11	37	40	40	46	25	15	29	1	1-46

given in Table 59. The last column gives the range, i.e., the highest and the lowest ranks assigned each subject. It will be seen that there is a marked disagreement among the interviewers. Applicant C, for instance, is placed first by one interviewer and fifty-seventh by another. The ratings of several other applicants show discrepancies of about this magnitude. It is to be remembered that these interviewers were sales managers with considerable experience in making such judgments; hence the extent of their disagreement in rating the same applicants is rather disquieting.

In another instance 6 sales managers interviewed 36 applicants for sales positions [28]. The results may be summarized in a word. In the case of 28 of the 36 applicants the managers disagreed as to whether the individual should be in the upper or the lower half of the group.

Another similar study was made in employing truck salesmen [30]. A want ad was inserted in the paper and on the basis of the letters received 12 applicants were selected. They were interviewed individually by 6 sales managers and also by a psychologist as to fitness for the position. There was fair agreement

¹⁶ From H. L. Hollingworth, *Judging Human Character*, by permission of D. Appleton-Century Company, New York.

among the interviewers as to the two best and the two worst candidates. In the other cases, however, the agreement was small indeed. The average deviation of the judges was a trifle over three places, and inasmuch as there were only 12 possible places this deviation was serious. After the estimates were pooled to secure a consensus, the reliability of any individual judge could be estimated by correlating his rating with the consensus. The correlations for the 7 judges were as follows: .12, .38, .01, .72, .58, .47, and .71. The last figure is the correlation for the psychologist; he did practically as well as any of the experienced sales managers. Two of these correlations, i.e., two interviewers, are fairly satisfactory, but we cannot be sure *a priori* that a given interviewer will be of this type. The results suggest the possibility of selecting interviewers on the basis of a tryout of this sort, with a computation of reliability looking toward building up a staff of interviewers who are really skilled. It is probable that the better interviewers are better because they employ more effective techniques, so the training of any interviewer—particularly an inferior one—in these specific techniques should bring about some improvement. Some such techniques will be discussed.

Improvement in Technique. One improvement is to limit the scope of the interview and to disregard aspects on which information can be secured in better ways. It would be foolish, for example, for the interviewer to estimate the applicant's intelligence when it could perfectly well be measured by a standard test. Anything that can be determined from tests, or from ratings by previous employers, or from valid objective items in the application blank should be obtained in that way; the interview should be limited mainly to data that can be obtained in no other way. Furthermore, it should be limited primarily to evaluating traits that can actually manifest themselves during the interview. These might include appearance, manner, likableness, emotional fitness for the job, disagreeable mannerisms. They would not include dependability, persistence, speed of reaction, or memory.

A second improvement in the procedure is to have a considerable number of interviewers. We have seen in other connections that if judgments are to be made, a better result can usually be obtained by pooling the estimates of several judges than is possible with an individual judge. If the applicant is interviewed by

several members of the staff and their judgments are pooled, there should theoretically be an increase in the value of the procedure.

Another desirable feature of the interview is the establishment of rapport at the outset. Just as in mental test procedure it is desirable to get the individual into the proper attitude so that he will cooperate and do his best, it is likewise necessary in the interview to get him into a proper attitude. This requires tact and skill on the part of the interviewer. The general attitude of the interviewer contributes in this respect. One who is in general sympathetic with the applicant and who can see the situation from the latter's point of view naturally will get closer to him in the course of the interview. In fact, some personnel departments consider this matter of the interviewer's having the applicant's point of view so important that they periodically send their employment men out job-hunting incognito. If they happen to secure a job, of course they do not report for it. But they make first-hand observations of the way employment situations are handled by other organizations and they see things from the standpoint of the applicant himself. After a week of it they have an intensive conference at which they all report their experiences. After having the door slammed in his face or standing around in an inhospitable waiting room or talking to a supercilious receptionist through a hole in a glass window, a man comes back to his own employment department keenly aware of the applicant's point of view. This facilitates the initial rapport and also the subsequent conduct of the interview.

Mention should be made of an interesting set of rules for the orientation of interviewers developed in the Western Electric Company [27, 272 ff.]. They will not be described in this connection because they apply fully as much to interviews conducted for improvement of industrial relations as to employment interviews. However, they stress such points as getting back of the manifest content of the interview and being alert for things the individual will not discuss without some encouragement.

A fourth feature is the use of crucial questions the value of which is known. The earlier consideration of items of personal history with reference to their validity will sometimes bring out items on which further information is desirable although they

are not a matter of record and this information must be secured from the interview. Job analysis may indicate the type of information that is most valuable in a specific case. Knowing what is needed, the interviewer can then seek the definite information that will be relevant. Care should be taken, however, that the questioning is not too cursory, but sufficiently flexible so that the applicant will reveal his own characteristics.

A final suggestion for improving the interview technique consists of having in mind during the interview certain specific traits which are to be observed. If, for instance, the interviewer is watching for such things as appearance, manner, energy, cooperativeness, and confidence, he should continually have these traits in mind during the conversation.

Occasionally the entire interview may be devoted largely to securing impressions regarding a particular trait. A case in point is the procedure for interviewing prospective prohibition officers with a view to determining their judgment and resourcefulness [22]. The interviewer presented to the applicant verbally a problem like the following: "Suppose I give you some information, and you are to question me in order to get clues. I am leaving the city tomorrow and suggest that you watch Mr. X in my neighborhood because I suspect that he and his brother-in-law have been violating the prohibition law. Now you proceed to question me to get further clues about the matter." Thereupon the applicant starts to question the interviewer. The latter has standardized items of information to release if the proper question is asked; otherwise, he does not release them. For example, if the applicant asks whether Mr. X has a car the standard answer is that he has an old Cadillac. If he asks where this tip was heard, the standard answer is that it was heard in a cigar store. If he asks where he thinks the liquor is obtained, the standard answer is Baltimore. The number of standard items of information that are brought out in this way gives a rough indication of the resourcefulness of the applicant.

Carrying out suggestions like the foregoing for improving interview techniques calls for some kind of an initial plan. This point is stressed by Smith [29, 271]. He suggests, for example, that initial plans may be made as to just what information the

interviewer wants to obtain, what information he wants to give the applicant, and what impressions he wishes to create.

Forms for Interviewers. The conduct of the interview along the lines just discussed may be facilitated by some type of printed form. Rather than let the interviewer proceed according to his own devices and vary the procedure from one interview to another, it is well for him to have before him certain set topics and items which he is supposed to investigate. Numerous such forms have been devised and are being used in various organizations; a few of them will be cited. Some suggest certain major topics, with subtopics or questions. In the following example, most of the questions are answered by either "yes" or "no," or by a check against one of three degrees of the characteristic such as good, average, or poor [18].

1. *Work History.* Questions under this include: Has he made progress in his work either with one company or from job to job? Has his past experience been helpful in the job? Does he appear to have liked his work?

2. *Aims in Life.* How much does he want the job? What features of the work appeal to him? How ambitious is he?

3. *Social Adjustment.* What is his attitude toward people in general? Has he ever demonstrated any leadership qualities? Is his social life adapted to proper work and study habits?

4. *Family and Domestic Situation.* Did his family influence him to form habits of industry? Did he work summers and after school hours? Is his wife working? What is the minimum income necessary to support him and dependents?

5. *Analysis of the Man's Motivation.* This item lists favorable and unfavorable motives that may have been involved, such as need to support dependents, looking for something easy, wants to try the job and see if he likes it, no need to earn money.

6. *Ratings on General Characteristics.* These may include appearance, voice, fluency, correctness of speech, and a considerable number of personality traits to be checked, such as aggressiveness, friendliness, forcefulness.

Some interview forms state the direct questions which the interviewer is supposed to ask [8]. For instance, work history includes the following: How did your previous employer treat you? What experiences of value did you get from each job? Can

you give any examples of success in your experience particularly in handling people? Other topics on which questions are formulated in this fashion are family history, social history, personal history.

A fairly recent interview form may be given completely [23]. It was developed for selecting salesmen for a company that sells soap and alkalizer to laundries, metal cleaners to the metal industries, textile cleaners to textile plants, and certain other heavy chemicals.

APPLICANT INTERVIEW FORM¹⁷

Name of Applicant Date
Address Interviewed by

FAMILY BACKGROUND

Are parents still living? Father's occupation? Mother's occupation? Father's educational background? Mother's? What sort of person was father? Does applicant depend upon him?

Are there brothers and sisters? Younger or older? Occupations of brothers and sisters? How successful is each? Was there competitive stimulus within the family and are there any signs of extreme dependence?

Regular chores around the house as a boy? Taught work habits and work attitudes? How get spending money? When start savings account? Learn value of the dollar in home? In general, was family background such that it is an asset or liability to applicant?

EDUCATIONAL HISTORY

Which subjects liked best? Which disliked most? In general, preference for science and math or English and social studies? Interested in practical or theoretical or artistic things?

Extracurricular activities. Sports, Dramatics, Fraternity, Social. Offices held. Work while in high school? In general, an average student, below or above average?

Major and Minor in College. (Same questions as above.)

Taken any special courses since leaving school? Speech. Selling. Busi-

¹⁷ After Otis.

ness subjects. What did applicant get out of them? Any courses in mind applicant would like to take?

PRESENT DOMESTIC AND ECONOMIC SITUATION

Married? Children? Any other dependents? Age, sex, schooling of children? Wife working? Any enthusiasm in talking about family?

What is wife's attitude toward (changing job) proposition? What is wife's attitude about being away from home? Will wife move to new territory willingly? Are wife's ideas important? Dependent on her?

Own home or rent? Mortgage? Current bills at stores? (Credit?) Any other indebtedness? Life insurance? Loans against it? Savings account now? Any reserve? Any other sources of income? Will living standard need to be adjusted?

What are principal hobbies or interests? Is life centered around the trivial such as sports, good times—or around getting ahead, building a future? Belong to many clubs? (Not too many scattered interests better.)

Does applicant appear emotionally upset for any reason? Health? Applicant has a home, family, income, etc., what else does he want? What working toward?

WORK HISTORY

Begin with first full-time job and work up to present application. Any jobs related to detergent selling? Any laundry? Any metal? Heavy Chemicals? Maintenance? Textile? Food and Beverage?

What type customer called on? Any distributor contacts? Work with distributor salesmen? Demonstration? Service? Any training experience?

How many jobs? How long on each job? Reasons for leaving each job (Friction, Dismissals, Good Judgment)? What job liked best? (Interest and enthusiasm about jobs?) Any strong loyalty or attachment to any employer? Has work experience been of value in job applied for? Approximate earnings on each job? (Past peak?)

Is he ambitious? Wide shifts in type of work? Can applicant handle technical aspects of job without training? How much training necessary?

INTENSIVE STUDY OF ONE OR TWO RECENT JOBS

Just what type of product was or is he handling? Does applicant know his company and his line thoroughly?

Describe a typical day's work. Just what did he do? How did he find his prospects? How did he make new contacts? What type sales approach did he use? ("Canned" talk, plan each approach, or trust to luck?) About how many calls a day? How long a day? Work regular hours—early and late?

How heavy were expenses? Keep any personal record system? (Systematic?) Has applicant testimonials or records of sales or earnings? Was story told enthusiastically? Did he stress money earned or sales made, rather than number of friends or quality of products?

REACTION TO PROPOSITION

Does applicant have any questions to ask about our proposition? (Note type of question.) What is general reaction to our proposition? What appeals particularly about it? (Making money. Quality product. Company reputation. Advancement opportunities. Future security.)

What is likely to be chief difficulty or handicap? (Is thinking hard-headed, planful, sound, wishful, emotional, overenthusiastic, confused?)

NOTE ALSO

Physical appearance. Vitality and energy. Businesslike interview. Well-organized and clear speech. Any obnoxious manners? (Egotism, indecision, self-critical, apologetic?)

What is applicant's attitude toward work, associates and superiors? Do hobbies and recreational activities, religious and political beliefs, closest friendship and social contacts indicate rounded, well-balanced life?

Will applicant be able to deal with customers? Can he handle distributor contacts? What will his effect be on distributor salesmen?

Rating Scale for Interviewer. The efforts to focus the attention of the interviewer upon certain characteristics which he is to observe may be facilitated by an actual rating scale. This has the additional advantage of reducing his judgments to quantita-

tive terms. This procedure should not be carried through without careful scrutiny, however. The rating scale may become merely a substitute for good judgment and by its appearance of objectivity lead to the erroneous assumption that an accurate evaluation of the subject has been made. But a well-constructed rating scale helps a conscientious interviewer to direct his attention to the important characteristics and evaluate them one at a time, just as an ordinary rating scale enables a rater to evaluate an acquaintance more systematically. The details of developing a rating scale will not be repeated, but it is possible to adapt the technique of the ordinary rating scale to the conditions of the interview. In order to make the usual ratings it is necessary to know the individual and to have observed him for some time. Some traits, however, may manifest themselves to a certain extent at first sight. Moreover, employment men often find it necessary to evaluate a man, at least to some extent, in a preliminary interview. This estimate may be better than nothing, and whatever can be done to increase its reliability is desirable. The actual traits to be estimated in any given case will depend on the local situation and the nature of the vocation and on whether they are of a kind that can be judged without long acquaintance.

The following is a portion of a typical man-to-man scale used for an employment interview. The interviewer is first provided with a rating scale blank on which he is to make up his master scale after the fashion described in Chapter XII.

INTERVIEWER'S RATING SCALE FOR EXECUTIVES

Make up a list of twenty-five or more executives whom you know very well. Include in this list some who rank very high, some who are intermediate, and some very low in traits such as appearance, energy, social attitude, tact, and initiative. Be sure that your preliminary list is representative.

Appearance and manner. Disregard every characteristic of the executive except the way he will impress people by his physical bearing, neatness, and facial expression. Consider whether he will be repulsive or whether he will fall somewhere between the extremes. Select from your list the man who ranks the highest in this respect and note his name on the first line, which is marked "Highest Mr." Then select the one who ranks lowest in appearance and manner and put him on the bottom line. Then, still considering only this same

factor, select a man on your list who falls midway between highest and lowest, indicating him on the middle line. Then determine one who ranks between the highest and the middle and another who ranks between the middle and the lowest, indicating their names on the appropriate lines.

HighestMr.	20
HighMr.	16
MiddleMr.	12
LowMr.	8
LowestMr.	4

Similar directions and blank spaces follow for energy, social attitude, tact, and initiative.

The interviewer at his leisure fills out a blank similar to the above. He now has his master scale by which he can evaluate the individual in the interview. The procedure then consists of having this master scale before him during the interview and actually comparing the applicant with the various men on the scale in the different traits there listed. If with reference to appearance, for instance, the applicant impresses him as similar to the first man listed on the master scale, he assigns him a score of 20. A separate "interviewer's rating blank" is of course provided for recording his judgments.

The interviewer may make these estimates during the conversation and record them as he forms them or he may hold them in his mind till the conclusion of the interview and note them shortly thereafter. In general it would seem better to make some notations during the interview. Certain acts of the applicant may indicate a marked presence or absence of some trait which would perhaps be forgotten before the end of the conference. The essential point, however, is that the interviewer has before him this concrete master scale with which he is comparing the applicant, man to man, while he is talking to him.

The method of rating by defined groups may likewise be used in this connection. A typical blank might read as follows:

INTERVIEWER'S RATING SCALE FOR EXECUTIVES

Imagine all the executives you know in the kind of position for which the applicant is to be considered divided into five classes of

equal size on the basis of each of the traits listed below. Have this blank before you and keep the classes in mind during the interview. Try to compare the applicant with these other groups and determine in which he should be located. Consider one trait at a time. If you cannot reach a decision regarding a certain one, pass on to the others and return to it later in the interview. When you have come to a decision regarding a certain trait, check in the appropriate column. You may grade as finely as you wish by placing the check toward the right or left of a column according as you consider that the applicant stands high or low in a given group.

	Low- est Fifth	Next Low- est Fifth	Mid- dle Fifth	Next High- est Fifth	High- est Fifth
<i>Appearance and manner.</i> How he will impress people by his physique, bearing, neatness and facial expression
<i>Energy.</i> Whether lazy and listless, gets things done, or is actual live wire
<i>Social attitude.</i> Whether meets people formally or halfway and informally
<i>Tact.</i> How he gets along with people
<i>Initiative.</i> Tendency to stick and get things done in the face of opposition

In similar fashion the graphic rating scale may be adapted to use during the interview. The following is typical of such a blank:

INTERVIEWER'S RATING SCALE FOR EXECUTIVES

During the interview have in mind the traits listed below. Try to observe the applicant as to the extent to which he possesses these various traits and indicate by a check mark on each line your judgment of the applicant. Be careful to judge each trait independently of the others. If it is feasible to make these judgments during the interview, do so, although it may be desirable to postpone some of them

until the end. After the interview, go over the results again immediately, for you may wish to make some slight revision.

Appearance and manner. Consider how he will impress persons by his physique, bearing, neatness and facial expression

Repulsive	Unimpressive	Satisfactory	Noticeable	Excites admiration
-----------	--------------	--------------	------------	--------------------

Energy. Consider the way he will presumably go at his work

Full of "pep," "live wire"	Active	Will get things done	Half-hearted	Lazy and listless
----------------------------	--------	----------------------	--------------	-------------------

Social attitude. Consider how he will act when meeting people in business way

Formal and constrained	Somewhat reserved	Will meet halfway	Cordial	Breezy and informal
------------------------	-------------------	-------------------	---------	---------------------

Tact. Consider his ability to get along harmoniously with others

Very tactful	Will seldom make a break	Will make occasional mistakes	Indiscreet	Antagonizing
--------------	--------------------------	-------------------------------	------------	--------------

Initiative. Consider his tendency to get things done in the face of obstacles

Meek	Irresolute	Moderate stick-to-it-iveness	Surmounts most obstacles	Very persistent
------	------------	------------------------------	--------------------------	-----------------

One other rating scale that more nearly approaches the checklist type may be mentioned [23]. It involves descriptions of the traits and a series of boxes in which the rater can check. This scale was developed for rating prospective salesmen in a detergent company and supplements the applicant interview form illustrated earlier.

RATING CHART FOR APPLICANTS¹⁸

1. KNOWLEDGE OF BUSINESS:

The applicant must know the laundry, metal, maintenance, textile, etc., business to sell products. If considerable training will be required, rate him low. If he can sell without further training, rate him high. Keep demonstrations, check-up service, and trouble-shooting in mind when judging. Check your judgment below.

Excellent ☐ Good ☐
No Training Required

Average ☐
Some Knowledge

Poor ☐ Very poor ☐
Little Practical Knowledge

2. MENTAL ABILITY:

Mental ability does not necessarily mean educational level. Will applicant be able to learn to sell our products? Is he bright and alert? Has he imagination? Will he show initiative and be helpful to the

¹⁸ After Otis.

distributor? Will he be able to plan intelligently? Check your judgment below.

Excellent ☐
Very superior

Good ☐
Superior

Average ☐
Ordinary

Poor ☐
Slow

Very poor ☐
Dull

3. BUSINESS KNOWLEDGE:

Is applicant a business man? Does he comprehend the problems in development, purchasing, manufacturing, accounting, and sales in a corporate business? Does he understand that our expense must be within our income, that there must be profit for stockholders? Does he possess enough business background and knowledge to represent us to distributors? Check your judgment below.

Excellent ☐ Good ☐
Considerable Business Knowledge

Average ☐

Poor ☐ Very poor ☐
Little Business Knowledge

4. PERSONALITY:

Consider how applicant impresses others by his manner, bearing, and tact. Does he smile readily? Does he talk too much or too little? Is he blunt? Too much self-importance, ego? How will he impress his customers? Will he work well with distributor salesmen? Will his personality be an asset in selling? Check your judgment below.

Excellent ☐ Good ☐
Outstanding Personality

Average ☐
Not Outstanding

Poor ☐ Very poor ☐
Personality Handicap

5. SPEECH:

Does applicant express his ideas clearly and easily? Speech need not be polished or grammatically perfect, but thoughts should be clearly and easily understood. Is he interesting? Enthusiastic about what he says? Speak to the point without wandering? Voice should be satisfactory enough not to detract from ideas. Check your judgment below.

Excellent ☐ Good ☐
Clear Expression

Average ☐
Satisfactory

Poor ☐ Very poor ☐
Questionable Speech

6. SALES ABILITY:

We are interested in sales ability as it affects the sale of our products. Has applicant convinced you that he has successfully sold for someone else in a difficult sales situation? Does he have a "nose for business"? Will he need help in difficult situations? Is he self-reliant? Is his best selling ahead of him or behind him? Does he have records or proof of his ability to sell? Check your judgment below.

Excellent ☐ Good ☐
Outstanding Sales Ability

Average ☐
Average Sales Ability

Poor ☐ Very poor ☐
Questionable Sales Ability

7. TEACHING ABILITY:

A salesman must work with distributor salesmen. Has applicant

ever demonstrated his ability to train? Was he successful? Can he present ideas clearly? Does he have the necessary teaching ability to train distributor salesmen to sell our products? Check your judgment below.

Excellent ☐ Good ☐
Outstanding Training Ability

Average ☐
Training Ability

Poor ☐ Very poor ☐
Questionable Training Ability

8. INDUSTRY:

Capacity for day-after-day plugging. On previous jobs did applicant work regular hours; make sufficient calls a day; dig up own prospects; work without supervision? Do you believe that he will be a hard and industrious worker? Check your judgment below.

Excellent ☐ Good ☐
Hard Worker

Average ☐
Work When Pushed

Poor ☐ Very poor ☐
Questionable Worker

9. PHYSICAL VITALITY:

Does the applicant impress you as having a surplus of physical energy? Is he vigorous, active, not sluggish? Is he physically large and impressive? If he is small in height or weight, he should make up for it in energy and health. Will he be able to travel, work hard? Check your judgment below.

Excellent ☐ Good ☐
Strong Energetic

Average ☐
Probably Satisfactory

Poor ☐ Very poor ☐
Slow Sluggish

10. ACQUAINTANCE WITH TERRITORY:

Does applicant know the territory for which he is being interviewed? Does he know customers in any territory? If not acquainted with prospects, rate low. If knowledge of territory and prospects is good, rate him high. Check your judgment below.

Excellent ☐ Good ☐
Well Acquainted with Territory

Average ☐

Poor ☐ Very poor ☐
Little Knowledge of Territory

11. SUMMARY JUDGMENT:

In making a final judgment about this man, keep in mind the situation which exists at the present time. You should know why the man is a good prospect as well as why you believe the man to be a poor prospect. Record your judgment and reasons below.

Excellent ☐ Good ☐
Unqualified recommendation

Average ☐ Poor ☐
Qualified Recommendation

Very poor ☐
Not recommended

Reasons for Summary Judgment.....
.....
.....

With blanks of this sort the interviewer can more adequately record his judgment of the applicant. The procedure for scoring

and weighting the items is exactly the same as that described in the discussion of rating scales.

This rating scale procedure probably represents the most valuable contribution that psychologists have made to date to the technique of conducting the interview. In the light of the general unreliability of estimates of traits and the danger that various factors will influence judgment, any efforts to put the interview on a more scientific basis are worth considering. The rating scale technique which has proved of some value in judging present employees may likewise contribute something to the improvement of methods of hiring individuals on the basis of a personal interview.

Validation of Scale. One study should be mentioned in which an interviewer's rating scale was validated [8]. The data were obtained from employees of a household finance organization. The items included work history, family and personal history, and a number of personality items, which had been arbitrarily weighted and combined into a total score. A comparison was made subsequently between employees who had remained on the job for a year and those who had been dismissed within a year. The total score on the scale used at the time of employment for the former averaged 24.6 and 21.3 for the latter, with a critical ratio of 4.5, indicating that the difference was significant. The rating scale apparently made some differentiation between those who stayed on the job and those who were not successful.

Interviewing Large Numbers. Interviewing procedures like those discussed above require a considerable amount of time on the part of the interviewer for each individual. If a company has a heavy load of interviewing, it may be necessary to organize it so that there is some preliminary sorting of applicants, a careful interview being given only to those who are possible employees or for whom an interview is desirable because of some matter of policy [19]. One organization that interviews several hundred applicants a day has the receptionist give a very brief interview and enter the data on an application form. Those who have some glaring shortcoming, such as lacking the educational requirements for the given job, receive no further interview and a note is made on the blank as to their shortcomings. Two full-time interviewers, one of each sex, conduct most of the necessary inter-

views. These include all "policy" applicants, such as friends of employees or stockholders. This procedure is supplemented with a notice in the reception room expressing interest in the applicant and indicating that the employment department will try to give him a frank opinion of the possibilities. A good many workers said that the greatest discouragement was not being interviewed at all. In times of severe unemployment when some applicants seem almost desperate, the interviewer may call up other concerns and try to make contacts even though he knows it is almost hopeless. The applicant hears one end of this conversation, and even a little encouragement such as this may go a long way toward maintaining his morale.

Other Functions of the Employment Interview. The foregoing discussion of the employment interview has dealt primarily with its fact-finding aspect. To be sure, its major function is to secure information about the individual's aptitude for the proposed job. However, certain other functions should not be minimized. For one thing, the interview should give, as well as obtain, information. Employment is not merely a process of selection—it should be a mutual process. The company is entitled, of course, to information about the applicant's qualifications, but the applicant likewise is entitled to information about the company and the proposed job. Many a man takes a job under false pretenses on the part of the company. He assumes that it is a steppingstone to other work, but discovers later that it is a blind alley. He is shocked to find that it has much greater hazards or is much dirtier or more irksome than he had anticipated. Or he has supposed that his duties would be mainly inside the building, but he discovers later that he has to go out on the road in cold weather. Probably it does not occur to him during the interview to inquire regarding these matters. However, the subsequent development at variance with his assumptions produces a dissatisfied employee, an outcome which could have been obviated by foresight on the part of the interviewer. The applicant is questioned, tested, rated, analyzed, and recommended, and he is entitled to some reciprocal information. In the interest of ulterior satisfaction and harmony the interviewer should put the cards on the table and tell the applicant about every aspect of the proposed job that is of possible significance. Even though the ap-

plicant at the moment merely wants *any* job with a pay envelope, he should nevertheless enter the job with his eyes open, knowing its disadvantages as well as its advantages.

In addition to giving information to the applicant the interviewer should strive to make a friend for the company. The traditional value of first impressions is important in this connection, and the first impression of the company which an applicant gets at short range is usually in the employment office. If the interviewer has the proper attitude and tries to interpret to him the company's policies and ideals, this may be the beginning of a permanent friendly relationship. An effort may be made to "sell" the company to the applicant. If this is successful and this attitude is firmly ingrained, it will often iron out some of the inevitable rough spots that arise in industrial relations, and the employee will stay with the company and remain loyal to it. Even if an applicant is rejected but goes away feeling that this would be a fine place to work and tells his friends about it, he constitutes an asset rather than a liability.

Other Types of Interviews. Passing mention should be made of other kinds of interviews which come within the sphere of the personnel man although not part of the employment program. The exit or termination interview is one. The employee who quits voluntarily is interviewed if he is willing. The principal interest naturally centers in why he is leaving, and frequently information obtained in such an interview will throw light on problems of industrial relations. If several people leave because they cannot stand their foreman, this points to a difficulty which calls for correction. Thus a termination interview may reveal sources of bad morale.

Another function of interviews is to give the employee a chance to express his opinions. Some organizations make it a policy to interview employees periodically, getting them to talk about their work and attempting to locate causes of poor morale which may be eliminated. If the interviewer agrees to keep things confidential and if he actually does so, the employees will gradually talk more frankly and much valuable information may be obtained in this way.

Another advantage is that an interview gives the employee an opportunity to get things "off his chest." A typical case is the

following. The interviewers in this particular company had been instructed to let the employees talk as long as they wanted to and get it all out of their system. One employee had developed privately a rather crack-brained system of philosophy but had never been able to induce anybody to listen to a complete exposition of it. In the course of the interview he mentioned something about his system of philosophy and the interviewer, catching his cue, remarked that he would be much interested in hearing about it. He did so for the rest of the morning and made arrangements for the employee to come back in the afternoon and finish the discussion. This was the first time in his life that this employee had found anyone who would listen to him while he outlined his system completely, and it made him a friend of the company for life. It was probably worth the interviewer's while to listen for several hours in the interest of the employee's lifelong loyalty. Other cases of this sort may be less striking; what is significant is the fact that morale is sometimes much better if employees have the opportunity of talking things through and getting everything "off their chest." In a broad personnel program this aspect of interviewing may be well worth considering.

SUMMARY

While the technique of mental tests is preferable for employment purposes to the use of less objective indications of vocational aptitude, there are situations in which these latter deserve consideration. They may sometimes prove a valuable supplement to tests. The more variables investigated, the greater the probability of finding some with high correlations with the criterion. Moreover, the best variables from the predictive standpoint are those which have low intercorrelations, i.e., involve discrete rather than overlapping factors; and it sometimes happens that the tests intercorrelate rather highly while the miscellaneous factors have low correlations with the tests. In this case the miscellaneous factors may well be embodied in the regression equation. In instances where test technique is not feasible, it is worth while to investigate these miscellaneous factors and determine their value so that they can be used more systematically than heretofore.

Academic record in school or college has some predictive value. The school curriculum itself is a selective process inasmuch as intelligent pupils progress more rapidly than the unintelligent. Statistical studies show that early marks in school are rather prognostic of later marks. Academic achievement, moreover, seems related to the type of success that leads to being listed in *Who's Who*, marks in technical school have been shown to bear some relation to subsequent salary, grades at West Point have been shown to be somewhat prognostic of military success, and in one instance salary in a large industrial organization was related to grades in college. The amount of education obtained by the applicant may show whether he has mastered certain fundamentals which he will need in his vocation, and his rate of progress as shown by age and grade at leaving school is an indirect indication of his general intellectual capacity. Efficiency in some types of work, such as clerical, has shown appreciable correlations with years of schooling. Each situation, however, must be investigated for itself because equivocal results have been found. Achievement in special school subjects, such as manual training, has rather obvious implications regarding aptitude for similar work. The choice of certain subjects in an elective curriculum gives some index of a person's interests and perhaps also of his abilities. However, school marks at best are distinctly inferior to tests when the latter are feasible. A three-hour test was far more predictive of the first two years' achievement in an engineering college than was the entire high school record.

It is sometimes possible from an individual's early proficiency in a given occupation to predict his subsequent success therein. With salesmen first-year production seemed a rather good index of production in following years. In a machine shop there proved to be a relation between the accidents encountered by an employee in successive quarters of the year. In some clerical jobs a considerable relation was found between efficiency in successive months, especially with reference to speed. It is necessary, however, to study the individual job because even in the same plant correlations between early and later success were high for some jobs and low for others.

The items on personal history or application blanks have been

analyzed as to their vocational significance. The technique consists of tabulating various biographical items for different occupational groups and noting which items are differential. The implications of height and weight for heavy muscular work are obvious. A more subtle problem is the relation of stature to selling ability. Salesmen as a whole are apparently appreciably larger than the average man, but within a given selling organization it is sometimes found that the largest men are the best producers and sometimes, and, perhaps more frequently, that the best salesmen are well above the average stature but not of extremely large proportions. The age of the employee has been found to be of significance among some kinds of clerical workers and especially salesmen. It appears that men of middle age are more effective in selling even when allowance is made for the effect of experience. A relation has also been found between age and turnover. In the early years there is the natural instability of youth seeking a vocational objective. Later, perhaps in the thirties, there is greater stability while the men are rearing families and buying homes. Then when domestic responsibilities lighten there frequently comes a search for one's ultimate vocation. This produces some instability until perhaps fifty, when interests have become fixed and a profitable change in employment is unlikely. The relation between stability and age does not hold in a period of unemployment, for then the younger workers, like everyone else, are satisfied with any kind of a job. The common notion that married employees are superior seems to have some statistical foundation. Other dependents also appear to afford an additional incentive for efficient work. A salesman with one or two children was apparently more effective than a man with more or less than this number. Previous experience may be of vocational significance from the standpoint of either its amount or its nature. It is not universally true, however, that the more experience the better, for other factors may bring a man with long experience to the employment office. The type of previous vocation may be of value in ascertaining the most profitable sources of supply for a given occupation. Job analyses may reveal patterns of qualifications for jobs which overlap. These various items of personal history make it possible to state

certain minimum qualifications in each respect. It is better to weight the items so that they can be combined into a differential score. This may be done arbitrarily or in some cases according to a regression equation. In such instances a better prediction can be made than by attempting to evaluate the items separately. Considerable investigation by these techniques has been carried on with insurance salesmen.

The letter of application has been studied with reference to its reliability and validity in indicating certain traits or general fitness for a position. Different judges disagree tremendously in estimating traits from a letter, and the same judge agrees with himself in subsequent estimates to only a fair degree. Estimates as to general fitness for a position give only fair correlations with estimates of some traits made by acquaintances; with other traits the correlations are negligible. However, the situation is improved by pooling the estimates of the judges regarding a given letter before comparing them with a criterion. This suggests that, if application letters are to be used at all, the most satisfactory procedure would be to have several members of the staff evaluate them independently and to combine their judgments. Generalizations as to character traits manifested by handwriting not only are without statistical foundation but have a certain amount of actual statistical refutation. Little significance should be attached to an applicant's own evaluation of his personality traits. Such an evaluation has been shown to be very unreliable and usually to involve too good an opinion of oneself.

Recommendations are often worthless because of the prejudice or carelessness of the writer. When this is not the case, note should be taken of the greater reliability of estimates of objective traits as compared with subjective. Moreover, the estimate depends on the conditions under which the recommender has observed the applicant. The testimonial and the letter written to the prospective employer at the applicant's request are of little value. The best form of recommendation is an answer to a specific inquiry from the prospective employer, because attention is thus centered on the particular information that will be of value. The technique of making inquiry may be improved somewhat by arranging a blank so that the recommender has merely

to choose certain alternative answers to questions or check in certain spaces. This saves the time of the one filling out the blank; moreover, he is more apt to do it seriously and in a less perfunctory manner. It also secures specific and unequivocal information.

The employment interview has certain shortcomings. The interviewer is prone to use personal physiognomic generalizations, to assume that a certain habit as manifested in one's appearance is general and will apply to his work on any job, and to misinterpret excitement in the interview situation as characteristic of the applicant elsewhere. When different members of the staff interview a group of applicants independently and then compare their estimates, the findings are rather disquieting. This was particularly true when sales managers interviewed applicants for selling positions. Suggested improvements in the interview technique are its limitation to items that cannot be evaluated objectively, the use of more interviewers, the establishment of rapport at the outset, and the use of crucial questions the significance of which has been established. The conduct of the interview along these lines may be facilitated by the use of printed forms. Finally, if the applicant is to be rated with reference to specific traits, it is possible to adapt the rating scale technique so that it can be used during the interview. The man-to-man scale, the method of defined groups, and the graphic rating scale or check list can all be used for certain traits that manifest themselves in a short space of time. This method makes it possible to rate the applicant on crucial points during the interview, to obviate the halo of general impression, and in general to obtain results that have greater reliability and validity.

In addition to obtaining information about the applicant, the interview serves two other important functions. It gives the applicant information about the proposed work so that he enters the job with his eyes open as to its nature, the working conditions, and the possibilities for advancement. Since there will thus be no discrepancy between his expectations and actual conditions, he will remain satisfied with his job. The interview also affords an opportunity to sell the company to the applicant and to make a friend. The interviewer should be not merely an examiner, but an instructor and a salesman.

REFERENCES

1. Becknell, H. E. Predicting Success of Y.M.C.A. Secretaries. *Personnel Journal*, 1928, 7, 172-175.
2. Bills, M. A. Selection of Casualty and Life Insurance Agents. *Journal of Applied Psychology*, 1941, 25, 6-10.
3. Blankenship, A. B., and Taylor, H. R. Prediction of Vocational Proficiency in Three Machine Operations. *Journal of Applied Psychology*, 1938, 22, 518-526.
4. Bridgman, D. S. Success in College and Business. *Personnel Journal*, 1930, 9, 1-19.
5. Goldsmith, D. The Use of the Personal History Blank as a Salesmanship Test. *Journal of Applied Psychology*, 1922, 6, 148-155.
6. Hollingworth, H. L. *Vocational Psychology*. New York, Appleton-Century, 1916, 308 pp.
7. Hollingworth, H. L. *Judging Human Character*. New York, Appleton-Century, 1923, 268 pp.
8. Hovland, C. I., and Wonderlic, E. F. Prediction of Industrial Success from a Standardized Interview. *Journal of Applied Psychology*, 1939, 23, 537-546.
9. Hull, C. L., and Montgomery, R. B. An Experimental Investigation of Certain Alleged Relations Between Character and Handwriting. *Psychological Review*, 1919, 26, 63-74.
10. Kenagy, H. G., and Yoakum, C. E. *The Selection and Training of Salesmen*. New York, McGraw-Hill, 1925, 380 pp.
11. Kitson, H. D. A Critical Age as a Factor in Labor Turnover. *Journal of Industrial Hygiene*, 1922, 4, 199-202.
12. Kitson, H. D. Height and Weight as Factors in Salesmanship. *Journal of Personnel Research*, 1922, 1, 289-294.
13. Kornhauser, A. W. A Statistical Study of a Group of Specialized Office Workers. *Journal of Personnel Research*, 1923, 2, 103-123.
14. Laird, D. A. *The Psychology of Selecting Men*. New York, McGraw-Hill, 1925, 274 pp.
15. Langlie, T. A., and Eldredge, A. Achievement in College and Later Life. *Personnel Journal*, 1931, 9, 450-454.
16. Lowell, A. L. College Studies and the Professional School. *Educational Review*, 1911, 42, 217-233.
17. Manson, G. E. What Can the Application Blank Tell? *Journal of Personnel Research*, 1925, 4, 73-99.
18. McMurray, R. N. How to Rate and Judge Personality Factors When You Interview Applicants for the Sales Force. *Sales Management*, 1937, 41, October 10, 11-14.

19. Nelson, C. M. Reception of Applicants for Employment. *Personnel Journal*, 1933, 12, 33-35.
20. Nicholson, S. W. Success in College and in After Life. *School and Society*, 1915, 2, 229-232.
21. Omwake, K. The Value of Photographs and Handwriting in Estimating Intelligence. *Public Personnel Studies*, 1925, 3, 2-15.
22. O'Rourke, L. J. Measuring Judgment and Resourcefulness. An Interview Technique. *Personnel Journal*, 1929, 7, 427-440.
23. Otis, J. L. Procedure for the Selection of Salesmen for a Detergent Company. *Journal of Applied Psychology*, 1941, 25, 30-40.
24. Phoenix Mutual Life Insurance Company. *Selecting the Successful Salesmen*. Hartford, Conn., 1937, 21 pp.
25. Poffenberger, A. T. School Achievement and Success in Life. *Journal of Applied Psychology*, 1925, 9, 22-28.
26. Poffenberger, A. T., and Vartanian, V. H. The Letter of Application in Vocational Selection. *Journal of Applied Psychology*, 1922, 6, 74-80.
27. Roethlisberger, F. J., and Dickson, W. J. *Management and the Worker*. Cambridge, Harvard University Press, 1939, 615 pp.
28. Scott, W. D. The Scientific Selection of Salesmen. *Advertising and Selling*, 1915, 25, 5-6, 94-96.
29. Smith, E. D. *Psychology for Executives*. New York, Harper, 1934, 311 pp.
30. Snow, A. J. An Experiment in the Validity of Judging Human Ability. *Journal of Applied Psychology*, 1924, 8, 339-346.
31. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
32. Super, D. E. Occupational Level and Job Satisfaction. *Journal of Applied Psychology*, 1939, 23, 547-563.
33. Swartz, G. O. Selecting Salesmen; Getting the Right Man the First Time. *Journal of Personnel Research*, 1924, 2, 457-459.
34. Thurstone, L. L. Mental Tests for Prospective Telegraphers. *Journal of Applied Psychology*, 1919, 3, 110-117.
35. Thurstone, L. L. A Standardized Test for Office Clerks. *Journal of Applied Psychology*, 1919, 3, 248-251.
36. Thurstone, L. L. Intelligence Tests for Engineering Students. *Engineering Education*, 1923, 13, 263-318.
37. Toops, H. A. The Selection of Graduate Assistants. *Personnel Journal*, 1928, 6, 457-472.
38. Young, P. T. Sex Differences in Handwriting. *Journal of Applied Psychology*, 1931, 15, 486-498.

Chapter XIV

TRADE TESTS

TRADE TESTS VS. TESTS OF INNATE CAPACITY

The distinction has already been made between tests of capacity or aptitude and tests of proficiency or achievement. In the case of the former we are concerned with innate aspects of the individual, certain potentialities which he possesses and which may be indicative of his subsequent success in the job—in short, with predicting what he will ultimately be able to do and with measuring a sample of his innate capacity that will make this prediction possible. Such tests are used, for instance, to determine whether he has the proper attention and reaction time to make a good tire-builder—a job that he has never tried before. In the case of tests of proficiency we are interested merely in particular acquired abilities or skills that he possesses now—for example, how good a carpenter or plumber he is when he enters the employment office. We do not attempt to prophesy; we merely try to determine present conditions. Inasmuch as proficiency tests are frequently used in hiring persons in the skilled trades, they are generally called trade tests. In a few cases where it was desirable to avoid the word “test” because of rapport, they have been called “work samples.” The conventional terminology will be used in the present discussion.

NEED FOR TRADE TESTS

The need for tests of proficiency arises in industry when hiring a person who is presumed to have a certain amount of trade experience. Machinists, carpenters, electricians, and the like apply for a job on the basis of their previous experience in their particular field. They frequently carry a journeyman's card or state

that they have served a certain length of time as an apprentice. Trade tests are designed to supplement this information. It is often undesirable to accept the applicant's own statement as to his proficiency or to take his card at its face value. This fact was vividly brought home to psychologists in 1918. Many military duties were of a specialized trade character and it was obviously desirable to assign to them soldiers who had functioned in a similar capacity in civil life. If a given unit contained a man who had previously been a barber and another who had previously been a plumber, there was obvious economy in giving them the same work to do in the Army rather than having the plumber cut hair and the barber mend leaks. Efforts were made to determine occupational status in interviewing recruits, but such interviews proved to be unsatisfactory. On the average, of the men who professed trade ability in an interview, 6 per cent actually proved to be experts, 24 per cent journeymen, and 40 per cent apprentices; 30 per cent were novices. In other words, approximately one-third of the recruits who claimed that they were carpenters could not drive a nail and one-third of the self-styled automobile mechanics did not know a spark plug from a carburetor. Hence it became imperative to develop some means for objectively determining a man's trade ability regardless of his own statement of his qualifications. The first extensive trade tests were developed in the Army, and subsequent developments have been much along the original lines.

This problem of trade tests, to be sure, is not of the magnitude of that previously discussed in connection with tests of innate capacity. The present trend in industry is toward a subdivision of labor so that a given worker performs only a relatively minor operation. Whereas formerly one person made the entire shoe, now one man cuts the sole, another cuts the upper, another stitches them together, and another puts on the heel, so that the trade of shoemaker is practically extinct. In one large concern, for instance, 4000 people work on the tools that are necessary for the automatic machines run by 15,000 others. Each man, moreover, works on only one machine so that there is no need as formerly for all-round toolmakers. However, many situations still require persons who actually have some trade proficiency. This is especially true of smaller concerns where the operations have

not been so minutely divided, but even large organizations need plumbers, carpenters, lathe operators, truck drivers, electric wiremen, and the like. Hence the test method of determining the actual trade ability of a prospective employee has considerable applicability.

REQUIREMENTS OF TRADE TESTS

Administration by Examiner with no Trade Knowledge. There are several requirements which a trade test must meet if it is adequately to fulfill its purpose in the practical situation. In the first place, it must be so constructed that it can be administered by an examiner who has little or no knowledge of the trade in question. This is necessary because of the frequent desirability of having the process of hiring centralized. A large employment office is often so organized as to do all the employing without consulting specific foremen or other members of the factory staff regarding individual applicants. Hence it would be almost impossible for the examiners to be completely familiar with all the trades involved. Even if the employment procedure were decentralized, there would be no guarantee that the foreman would administer all the tests in the same fashion. It is obviously fundamental from the scientific standpoint to give every applicant exactly the same test procedure.

Score Independent of Examiner's Judgment. In the second place, the tests should be so constructed that they yield a rating independent of the examiner's judgment. The score should be entirely objective and quantitative. This point is related to the preceding. If it were necessary to rate an iron hook made by an alleged journeyman blacksmith as excellent, good, average, fair, or poor, it is probable that there would be marked disagreement between raters. One of them might note whether the ring at one end was perfectly round, another might be more concerned with the shape of the point, while a third might note specially whether the general dimensions conformed to specifications. If the hook happened to be well made in one of these respects but not in the others, the applicant's rating would depend largely on who rated his test. It is possible, however, to devise tests in such a form that they can be administered by a person with no trade knowl-

edge and nevertheless yield the same result for a given applicant regardless of who examines and rates him.

PRINCIPLES ON WHICH TRADE TESTS ARE BASED

There are two principles according to which test material may be constructed. An individual who is successful in a trade possesses, on the one hand, a certain amount of skill and, on the other, a certain amount of information. A machinist, for instance, is able to set the chisel and operate the feeds on a lathe. He also has certain information about a lathe and can tell the difference between the headstock and the tailstock. In attempting to determine whether he has had experience in lathe work there are thus two possible avenues of approach. We may ascertain through some standard performance just how well he can manipulate the parts or we may find out how much information about the machinery and materials he has acquired. The information type of test is especially valuable in its negative aspect, i.e., in eliminating those who make false claims to trade ability. If a man lacks the information, obviously he has had little contact with the trade. If he has the information we still cannot be entirely certain about his skill. In general, however, a skilled tradesman will be able to give a good account of himself either in actual performance or in answering questions pertaining to his work.

KINDS OF TRADE TESTS

Oral. The different kinds of trade tests that have been used fall into four general classes: oral, picture, written, and performance tests. Illustrations of each type will be given later. In the oral test, which is still used quite extensively, the examiner is provided with a blank that contains the questions, space for the applicant's answer, the correct answer, and the credit to be given for each answer. The examiner reads the questions to the applicant and writes in the latter's answers. The questions deal with tools, materials, processes, and other information concerning a trade that a worker in that trade would be apt to have at his command.

Picture. In the picture trade test the applicant is questioned regarding the details in pictures of machinery or tools used in the trade. For administering the test two folders are usually pro-

vided, one for the examiner and one for the applicant. The latter contains the pictures numbered in sequence. The examiner's folder contains the questions similarly numbered, as well as the answers and proper credit for each.

Written. The written trade test is somewhat similar to the oral except that it is designed for group administration. This necessitates making it sufficiently fool-proof so that the subject can respond adequately by writing or making check marks. The multiple choice form of response employed in capacity tests is generally used in written trade tests.

Performance. In the performance test the applicant goes through some standardized typical operation which can be scored on the basis of how he does it or by evaluating the finished product. In scoring, concern is not, as in the other types of trade test, with whether the subject gives or fails to give a certain answer. It is rather a matter of a complex operation or product that must be evaluated. In the first place, the process that the subject uses in taking the test may be considered by itself. In a performance test for a truck driver what would be observed primarily is the way he handles the truck in going through prescribed maneuvers. In the second place, the product itself may be rated. A blacksmith may be required to reproduce an iron hook like a sample and then be graded on the finished product according to how well he actually makes the reproduction. In the third place, the time consumed in making the product may be the important consideration.

In general practice these three methods are seldom discrete; two or three of them are combined. For instance, a process-time test may be used in which the man is required to change the set-up of a lathe in order to do a different job; he is scored according to the steps taken in making the change and also the length of time consumed. A critical score may then be established on the basis of both performance and time. Similarly, a product-time test may be used. A typist is given a piece to copy, the finished product is evaluated, and the time taken to complete the work is noted. It is possible also to use a process-product-time test in which all three items are considered. If, for instance, a garage mechanic is given a radiator to repair, his score can be based on the method he uses, the completed job, and the time consumed.

The type of performance test most generally used is probably the product-time test. Its advantage over the process test lies in the fact that it can be scored at leisure and without any expert knowledge on the part of the scorer. There are situations, however, in which the process-time test is more satisfactory. In most cases, at any rate, the time is taken into consideration.

Relative Advantages and Disadvantages. Each of these types of test has certain merits. The oral test has the advantages that characterize any individual as contrasted with any group test. There is the possibility that the applicant will misunderstand some trivial point, but this can be detected immediately by the examiner. If the subject has any difficulty in making himself clear, he can do so more effectively in conversation. Moreover, he may manifest certain reactions extrinsic to the test, such as emotional instability, that will be of vocational significance; this "clinical aspect" is present in the oral procedure, but missing in the group test. Finally, the fact that rapport is usually good is one important reason why the oral test is still widely used.

The picture test is usually conducted orally; hence it has the foregoing advantages. In addition it has certain other desirable features. It approaches more closely to the actual job situation. Looking at a picture of a machine tool gives a more tangible idea than merely talking about it. It gives the applicant more confidence in the test because it seems more concrete and apparently more practical. It admits, too, of more intricate questions because questions can be asked about the more minute parts that can be lettered on a picture but that would be difficult to describe adequately in an oral test. It is also possible that a picture will help the subject to recall further facts because it will be associated with various things in his work and thus help him orient himself.

On the other hand, there are disadvantages in the picture test. Such a test is somewhat more difficult to construct and it is more expensive inasmuch as it involves printing pictures. There is also the danger that the picture will be slightly atypical of the machine with which the worker is familiar. A man who is used to a lathe driven by an independent motor may be a trifle confused when shown a picture of a lathe driven by a belt from a main power line. This slight confusion may be enough to mislead him

on lathe questions. Finally, if there are several questions about one picture and the applicant fails to recognize the picture at all, he is unduly penalized because he will fail on all the questions dealing with it.

While the written test lacks the clinical advantages of the oral, it makes for much more rapid testing just as is the case with any group test. The multiple choice form of response likewise has the usual advantages. In the first place, the subject does not have to phrase his own answer. Certain individuals with poor ability in grammar might make a rather unfavorable showing although they were good workers. In the second place, there is no doubt as to the correctness or incorrectness of an answer. The person scoring the blank does not have to judge or subjectively evaluate an item. To be sure, the questions may be so selected that only a single correct answer seems possible, but even then there is always the possibility of an unsuspected answer that will give some indication of familiarity with the trade. When it is simply a matter of selecting one of several alternatives that are sufficiently discrete, there can be no question as to whether or not the subject deserves credit. In the third place, this type of test can be scored by anyone even though he is unfamiliar with the occupation in question. Finally, the multiple choice form makes possible more rapid scoring by the use of a stencil which can be aligned over the blank and which enables incorrect answers to be readily located.

The performance test has the advantage of dealing with actual trade skill rather than with information. It is possible for a person to work at a trade and pick up the information without acquiring the requisite skill. This is probably much more likely than the opposite tendency to acquire the skill without the information. The skill test is a more direct approach to the ability in question. In typewriting, for instance, it is more important to operate the machine effectively than to know the names of the different parts or the adjustments or the difference between various kinds of machines. The performance test is usually more difficult to arrange than the oral or written test. It requires a certain amount of equipment and often materials that are used up in the process of taking the test. If sufficiently complicated so that fairly elaborate equipment is required, it must be given as an

individual test. Similarly, if it is scored as a process test, it requires one examiner per subject, and consequently if given on a large scale its administration is more expensive than in cases where the group method is possible.

METHODS OF DEVELOPING AND STANDARDIZING TRADE TESTS

Differ from Methods of Developing Capacity Tests. The methods of developing and standardizing or calibrating trade tests are rather similar for all four kinds. For purposes of illustration the method will be presented in detail only for the written test, but the technique to be described is typical of methods of evaluating the other tests as well. A somewhat different approach to these problems is necessary from that employed with tests of innate capacity. In that case we were concerned with a group of separate tests such as those for attention, memory, or decision, each composed of many items. The total score for each test was then correlated with the criterion in order to determine the relative importance of the tests and combine them ultimately into a single score. In a trade test, however, all the items may be of one sort—for example, items dealing with trade information—so that it is not possible to evaluate separately a number of different tests, each composed of many items. Consequently, it is necessary to analyze the individual items and look for those that are most differential of trade ability. This procedure does not lend itself readily to the computation of correlation coefficients. It is more like the item analysis employed with measurements of interest (p. 328). The usual practice is to take a few groups of subjects differing in known trade ability and see which particular items or questions differentiate these groups.

Securing the Criterion. A criterion that is frequently used in this procedure involves grouping the applicants according to the ordinary trade classification of novice, apprentice, journeyman, and expert. Most trades have more or less definite standards of their own regarding these classes. The following definitions have been used extensively in some projects. An expert is defined as “a man with a high degree of trade ability qualifying him for assignment requiring superior workmanship”; a journeyman, as “a man with enough trade ability to qualify him for assignment

to work which must be done quickly and well"; an apprentice, as "a beginner or man with only enough trade ability to make him useful as a member of a group under supervision, not qualified to work without supervision, or where speed and accuracy are prime factors"; and a novice, as "a man with no trade ability, or so little that he should not be considered when making assignments." When the subjects have been so classified, it is then possible to determine, for each individual question, what percentage of the experts answer it correctly, what percentage of the journeymen, what percentage of the apprentices, and what percentage of the novices. If these percentages decrease in the above order there is some indication that this item is differential of the trade ability in question. This method will be described in detail below.

There is nothing mandatory about this fourfold criterion of novice, apprentice, journeyman, expert. In much of the work done by the Occupational Research Program of the U.S. Employment Service the criterion consists of three types of individuals, A, B, and C [2, 35]. Class A constitutes experts, individuals who are considered by their superiors to be thoroughly skilled in the occupation. They have a minimum of four years of paid experience as experts, although occasionally the substitution of someone a little short of the four years is permissible providing a special statement is made regarding him. Class B includes apprentices and helpers, and others who are not considered by their superiors to be thoroughly skilled in the trade. Class C consists of persons in related occupations, rather than the usual miscellaneous novices. For example, in developing tests for painters Class C included carpenters, paperhangers, sheet metal workers, plasterers, and glaziers. These workers would have some familiarity with a painter's work through having perhaps worked on the same house construction job but would not be skilled painters themselves. Applicants for a position who claim trade ability and do not have it are quite apt to be persons from related occupations like those just cited. The important point, then, is to differentiate between a skilled painter and a paperhanger who misrepresents himself as a painter.

In some cases a more detailed criterion has been used. At one municipal employment center the workers were classified on the

basis of ten degrees of ability: (1) novice, (2) good novice or handyman, (3) new or poor apprentice, (4) average apprentice, (5) excellent apprentice or poor journeyman, (6) poor journeyman, (7) average journeyman, (8) common journeyman, (9) expert, (10) thoroughly competent expert.

The procedure of selecting and standardizing trade questions in this fashion will be illustrated by an information test for lathe operators. This particular test was devised with the multiple choice type of response, which adapted it to the group method. However, the technique used in developing it is equally applicable to the other types of information test and to some extent to the performance test as well.

Preliminary Selection of Items. The first step in the development of such a trade test is to make a preliminary selection of items. If the psychologist engaged in this development knows little or nothing about the trade in question, it is advisable for him either to consult trade journals or to discuss the matter with foremen and possibly expert workers before devising the test items. In most instances the latter procedure is followed. If the nature of the project is made clear and the principle of finding questions which the good worker can answer and the poor one cannot is explained, the average foreman will see what is wanted, and he can then be asked to suggest a preliminary set of questions. Careful observation of the job will suggest other questions.

These preliminary questions must be worked over carefully before they will be in satisfactory final form. Many of those originally devised will be found to be indefinite or equivocal. It is not feasible to give the foreman a course on the construction of trade test questions. It is better to take his initial attempt and show him where improvement can be made. It may be advisable to give the questions individually to a few workmen and ask for their comments.

The following typical original questions were obtained from a foreman who supervised engine lathe operators. The general program was explained to him and he was requested to submit questions which might be used to determine whether a workman had the requisite information regarding the trade.

1. What is an engine lathe?
2. What is a lathe dog?
3. How fast should a belt run?
4. What is the most vital feature of a lathe?
5. What is the outside diameter of 1" pipe?
6. What is the correct angle for lathe centers?
7. What is meant by the pitch of gears?

Cursory analysis will reveal the ambiguity or equivocal character of some of these questions. Number 1 is far too indefinite. It might require anything from an elaborate definition and description to a brief statement as to what a lathe does. Question 2 is somewhat similar in that nothing is stated as to whether the information desired concerns the shape of the lathe dog or its function. Question 3 does not indicate in what unit the answer should be given. Question 4 implies some indefinite standard as to what is meant by vital feature. Question 5 is misleading in that one does not know how exact an answer is necessary. The last two questions are somewhat more specific and definite.

Revision of Preliminary Items. All the questions were carefully reviewed in this fashion and the findings transmitted personally to the foreman who had originally submitted them. Then in conference with him the above questions were revised as follows:

2. A LATHE DOG is used to TIGHTEN THE CHUCK; DRIVE THE WORK; LOCATE CENTER; CROSS FEED.
3. How many feet per minute should belts travel for the best results? 2000; 4000; 6000; 8000.
4. The most vital feature of a lathe is CARRIAGE; ALIGNMENT OF STOCKS; BACK GEARS; TOOL POSTS.
5. The approximate outside diameter of 1" pipe is 1 1/4 inches; 1 5/16 inches; 1 3/8 inches; 1 7/16 inches.
6. What is the correct angle for lathe centers? 60°; 45°; 55°; 70°.
7. PITCH OF GEARS means SHAPE OF TEETH; WIDTH OF TEETH; NUMBER OF TEETH PER INCH; ANGLE OF GEAR TO SHAFT.

The difference between the revised and the original questions is obvious. Question 1 is dropped as being entirely too indefinite. Question 2 restricts consideration to the function of the lathe dog by providing alternative functions. Question 3 indicates the units in which the answer is desired and is so arranged that there is

a wide difference between possible alternatives. Whereas, if left to his own discretion, the workman might debate between 3950 and 4000 feet, in the present form of this question he would have little hesitation, if he knew anything about the operation, in deciding between 2000 and 4000. Question 4 becomes much more specific by precluding the possibility of any very general comment. Question 5 states the terms in which the answer is desired, rather than leaving the subject to determine for himself how fine to make his estimate. The last two questions remain intact, but the multiple choice answers are added for the sake of uniformity. Similar revision was made of the other questions originally submitted.

After the questions had been recast into this form, they were submitted to other foremen for further suggestions and criticisms. They were also given to a few workmen who were not to be included in the final study, in order to determine whether any ambiguities had been overlooked. A few defects were discovered in this way and appropriate correction was made. The result of this preliminary selection and analysis was a set of questions ready for final selection. In the present case 60 questions similar to those in the above illustration were retained.

When seeking trade test questions it may be helpful to think in terms of a standard list of aspects of the trade concerned rather than in random fashion. The following eight classes cover the ground fairly well [2, 45]:

1. *Definition*, e.g., "What is a shore?" (carpenter)
2. *Limitation*, that is, some limitation or modification of the materials or tools or machine. "What is the smallest number of cuts necessary to mill a six-sided nut?" (machinist)
3. *Use*, generally involving a specific question as to what tool or material would be used for a particular purpose. "What device is used to test the specific gravity of the electrolyte?" (electrician)
4. *Procedure*, usually concerned with what to do for a specific purpose. "What do you do to the outside of a manhole?" (bricklayer)
5. *Location*, e.g., "From what part of the animal are pork chops usually cut?" (butcher)

6. *Names*, such as the names of tools, machines, or processes. "What is half a brick called?" (bricklayer)

7. *Purpose*, usually introduced by the word "why." "Why is tin added to a brass mixture?" (foundryman)

8. *Numbers*. "How many jaws are there in a universal chuck?" (machinist)

These eight items should be helpful in formulating new trade test questions. If one thinks over things in the trade that have a specific location or a particular use or thinks in terms of "why" questions he will probably have a considerably larger list of items from which to make the final selection.

When developing trade tests that are to be used on a nationwide basis it is necessary to be on guard against colloquialisms. If the answer to a question is "kerosene," in many places the response is more apt to be "coal oil." A question about the type of asphalt used on a flat roof was answered correctly west of the Mississippi but more and more incorrectly the farther east the verification process proceeded. Problems like this would not arise in the ordinary personnel organization unless tests were being developed for use in branches all over the country.

Final Selection of Items. Comparison with the criterion must be the basis for the final selection of the items. The questions must be given to groups of workers with varying degrees of ability to determine on which questions the best workers make the highest scores. It is sometimes possible to obtain the criterion from the men's trade union ratings. In other cases the foremen may estimate the men according to the conventional classes of trade ability. In some instances production figures can be obtained, but in many of the trades the work is too complex, the individual workmen perform widely varied operations, and they produce a single complex object rather than a definite number of "pieces." For the lathe operators discussed above, the foreman's judgment alone was available. He classified the men into three groups comprising experts, journeymen, and apprentices. In addition, the questions were given to individuals entirely outside the industry who might be considered novices.

The questions were presented on a mimeographed blank in the usual way, with directions and illustrations explaining how

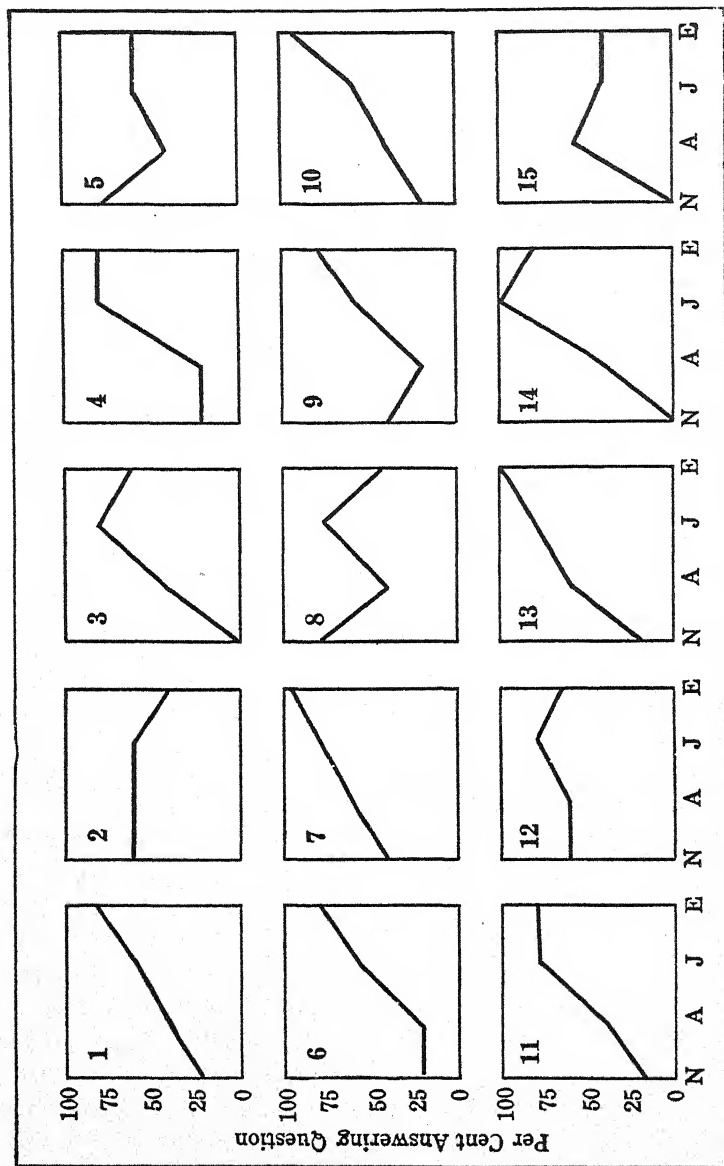


FIG. 9. SELECTION OF TRADE TEST ITEMS

to indicate one of the four alternative answers as correct. No particular time limit was set for the test. It was given to the lathe operators and novices in small groups at their convenience. The results for the first 15 questions are shown in Fig. 9. Each little diagram gives the results for one question. The four classes of trade ability are laid off along the base line and the percentage of each class answering the question is indicated by the distance above that point. The diagram for Question 1, for instance, shows that this question was answered correctly by 20 per cent of the novices, 40 per cent of the apprentices, 60 per cent of the journeymen, and 80 per cent of the experts. This question is very satisfactory. As we go from novices through the other classes to experts, there is a steady increase in the proportion who answer the question. This ideal curve is rarely attained in actual practice, but any question whose graph approaches the ideal within reasonable limits may be considered satisfactory. A glance at the other curves in the figure shows that some questions are manifestly worthless and others give at least some degree of differentiation.

There are various types of differentiation. Question 4 differentiates rather sharply the experts and journeymen from the apprentices and novices, although it does not differentiate between the experts and journeymen or between the apprentices and novices. Question 3 separates the novices from the other three classes without indicating consistent differences among these three. After surveying charts like those in the figure for each of the 60 questions, it was possible to select a limited number which seemed rather differential of the trade ability. Of those in Fig. 9 the following were retained: 1, 3, 4, 6, 7, 9, 10, 11, 13, 14. These and 30 others constituted the final set of 40 questions that comprised the trade test. This, then, completed the selection of items.

Instead of constructing a graph for each question, decisions can be made on the basis of the mere percentages. An additional refinement of method is to consider the statistical significance of the differences between percentages, i.e., divide each difference by its standard deviation to obtain the critical ratio. (Cf. p. 328.) This method has been followed by the U.S. Employment Service

in developing many trade tests. Their criterion, it will be recalled, consisted of A, experts; B, apprentices and helpers; C, workers in related occupations. Significant differences between the percentages of these classes answering a question were sought. More stress was laid on differences between the A and B groups because in the situations where the tests were to be used this differentiation was more important than that between the B and C groups.

Calibration of Final Set of Items. One step remains, namely, to calibrate the set of questions finally selected and set critical scores. Suppose that a prospective employee has been given the 40 questions and makes a certain score; it becomes necessary to interpret this score with a view to ascertaining his presumable trade status. It is desired to determine what degree of trade proficiency may be expected from a man who scores 10 points or 20 points. The procedure for determining a critical score for a trade test is analogous to that discussed previously. It is desired to obtain some score above which there is a strong probability of the individual's being an expert and below which the chances are that he is a journeyman. It is also desired in similar fashion to draw the line between journeyman and apprentice and between apprentice and novice.

A graphic method similar to that described for aptitude tests (p. 249) is feasible. This was used with the data for engine lathe operators discussed above. The calibration is shown in Fig. 10. The possible points of test score are laid off along the base line. The figure is divided into four sections, one above the other. The topmost represents the experts, the next the journeymen, and so on as indicated by the letters at the left. Each square represents one man and is located in the proper trade class and directly above his score on the base line. It is obvious that the squares representing experts appear farther to the right than those representing novices. The problem now is to draw a vertical line that will make the best division between the journeymen and the experts. If, for instance, this line is drawn between 24 and 25, all the experts will be to the right of the line and all but one of the journeymen to the left. If it is drawn between 26 and 27, all the journeymen and only one expert fall below this point. Either of these critical scores seems satisfactory, inas-

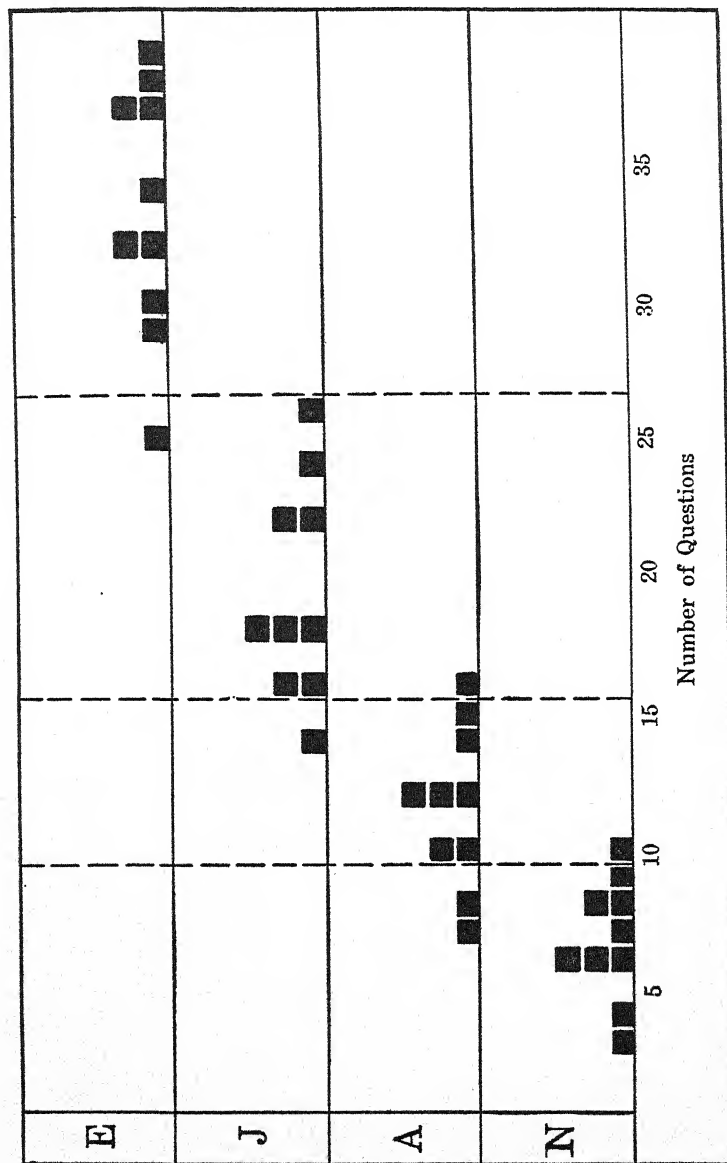


FIG. 10. CALIBRATION OF TRADE TEST ITEMS

much as only one man is displaced. It can then be said that if a man scores 27 or more points, the likelihood is that he is in the expert class. In similar fashion a line between 15 and 16 will clearly separate the apprentices from the journeymen with the minimum overlapping. With reference to the novices and apprentices the separation is a little less sharp, but the best point seems to be between 9 and 10.

In such calibration procedure it is usually not possible to make an abrupt separation between two classes, and the experimenter will have to use good judgment in determining the best place for the line. The essential point is that by drawing the line at the proper point most of those to the right will be in a trade class superior to those at the left. The points at which these lines are drawn thus constitute the critical scores and are the final figures that are desired in order to interpret the scores of any applicants who subsequently may take the trade test. In the present illustration they may be stated in the following form for convenient reference:

27 to 40	Expert
16 to 26	Journeyman
10 to 15	Apprentice
0 to 9	Novice

Another procedure for setting the critical score may be illustrated by hypothetical data in Table 60 [2, 215]. The criterion consists of the A, B, and C classes described above (p. 485). The first column gives the scores. The next column gives the percentage of the subjects in Class A—experts—who made a score equal to or higher than the one indicated. For example, 41 per cent of the subjects in Class A scored 13 points or better, whereas 74 per cent of them scored 12 points or better. The next column gives similar data for Class B—the helpers and apprentices, and the next column for Class C—the persons in related occupations. The final column gives an index which differentiates Class A from the others, namely, $2A - (B + C)$. As we run down the column we see that this index rises to the maximum for a score of 8. This point is taken as the critical score and applicants who receive a score of 8 or more on this particular set of 15 ques-

tions would be considered expert. A similar procedure could be used to differentiate Class C from the others, the formula being $2C - (A + B)$.

TABLE 60. SETTING CRITICAL SCORE IN TRADE TEST

Score	Per Cent of Persons at or Above the Indicated Score			2A - (B+C)
	A	B	C	
15	3			6
14	15			30
13	41			82
12	74			148
11	86	4		168
10	94	12		176
9	97	12	3	179
8	100	16	3	181
7	100	24	3	173
6	100	28	6	166
5	100	52	6	142
4	100	72	9	119
3	100	80	17	103
2	100	84	33	83
1	100	100	57	43
0	100	100	100	0

EXAMPLES OF TRADE TESTS

It now remains to illustrate the different varieties of trade tests. The foregoing discussion dealt only with the written test, but the oral, the picture, and to some extent the performance tests are generally similar in their method of development and calibration. However, inasmuch as their content varies somewhat, examples of each will be given.

Oral Trade Test. It is not necessary in the present connection to give a complete set of questions for any particular trade test, for complete forms for many such tests are available elsewhere [1, 3]. In what follows only a few items for each test will be included by way of illustration. Each question is given in the form

in which it is asked, followed by the correct answer. In a few instances two or more answers are allowable.

Painter

1. What do you do to knots and sappy places before painting?
Shellac.
2. When is puttying done on new woodwork?
After priming (first coat).
3. What is the brightest yellow used?
Chrome.
4. What do you use to bleach an exposed oak door before refinishing?
Oxalic acid.
5. What device is used for working just outside of a single window on a high building?
Jack.

etc.

Bricklayer

1. Where do you use radial or round-face brick?
In arches.
2. What is used in the middle of a long wall to keep the line level?
Twig (twigger) (twigging) (tingle).
3. What is a brick called when set on end?
Soldier.
4. What is a bond called when a header and stretcher are laid in the same course?
Flemish.
5. What is the course called from which an arch starts?
Spring (springer) (springing course).
Skewback.

etc.

Carpenter, Finish

1. What kind of a bit is used to drill a 2-inch hole?
Expansion bit.
2. When driving into hard wood, what do you put on a finish nail to make it go in easier?
Soap or paraffin.
3. What side of a door is usually fitted first?
Hinge side.
4. What is used to secure the nailing of baseboards in a brick wall?
Base blocks.

5. What size finish nail should be used on a door jamb?

Tenpenny.

etc.

Cook, General

1. What do you use to clear boiled coffee?

Egg shell.

2. How would you cook lamb chops or steaks for children?

Broil.

3. How hot an oven should you have for biscuits?

450 degrees.

4. Do you start soup in hot or cold water?

Cold.

5. What do you put on fried sweet potatoes to make them brown?

Sugar.

etc.

Auto Mechanic, General

1. What joint is there between the differential and the transmission?

Universal.

2. What regulates the height of gasoline in the carburetor?

Float or float valve.

3. What are the marks on the flywheel used for?

Timing.

4. If a cylinder is scored from overheating what repairs are necessary to put it in good condition?

Rebore and regrind.

5. What tool would you use in trueing up bearings?

Scraper.

etc.

Picture Trade Test. The method of developing the picture trade test is essentially similar to that used for the oral. Various pictures and questions based thereon are selected and tried out to determine whether the skilled workmen on the average answer them more satisfactorily than do the unskilled. A few typical items from a number of picture trade tests will be described.

Carpenter

The test includes a series of pictures of tools, the question for each one being "What do you call that?" Pictures are shown of such things as a jack plane, spoke shave, saw clamp, draw knife,

ripping chisel, scraper, and miter. There is also a picture of a flight of steps with letters indicating the rise, the tread, and the nosing, and the applicant is asked to name the different parts that are lettered. A picture of a roof is shown with the valley and the ridge indicated. The applicant must name them.

Storage Battery Electrician

Pictures of four battery units are shown connected in different ways—in series, in parallel, and in a combination of series and parallel. The applicant is asked to tell how many volts will be obtained under these different conditions. There are also pictures of plates from various kinds of batteries which the applicant must identify. There are pictures of damaged plates with questions as to what might have caused that particular kind of damage. A charging system is illustrated and the applicant is required to point out the fuse and the resistance switch and to state what kind of current would be used in the circuit.

Machinist

A test for machinists involves pictures of different kinds of chucks—4-jaw, 3-jaw, and drill—which the applicant must name. There is a picture of a turret lathe with a question as to what kind of lathe it is. A vernier scale is set at a certain figure and the applicant is required to read it. He also has to name from the pictures various types of cutting tools and a number of different kinds of gauges.

Written Trade Test. The method of developing the written trade test was outlined in the discussion of methods (*supra*). Each item of information regarding the work includes a question and several alternative answers, the correct one of which is to be checked. Similar items for a few other trades will be given.

Bricklayer

1. Half of a brick is called: CHUNK; BLOCK; HEEL; BAT.
2. Fire bricks are laid in: CONCRETE; CEMENT; FIRE CLAY; MORTAR.
3. The top course of stone on a wall is called: COPING; BONDSTONE; CLIPCOURSE; CAPSTONE.
4. Before plumbing up a corner you should lay: THREE COURSES; SIX COURSES; NINE COURSES; TWELVE COURSES.

5. A fire stop around a flue is formed by a COPING; SKEWBACK; CORBEL; INDENT.
6. To keep the line level in the middle of a long wall you use: LEVEL; PLUMB LINE; SQUARE; TRIGGER.

Time Clerk

A written test for time clerks involves items such as the following: Two sheets of numbers are to be added quickly—numbers like $8\frac{1}{2}$, $8\frac{3}{4}$, $11\frac{1}{4}$ —the type ordinarily added by a time clerk in computing hours and fractions thereof. There are likewise two sheets for subtracting times such as the time between 7:30 and 11:15 A.M.—another type of computation performed repeatedly by time clerks.

Student Engineer

A trade test has been devised by one of the electrical concerns for selecting student engineers. This is essentially an information test dealing with data which these students should have learned before applying for such a position. Three types of items are intermixed throughout the test. The first involves lists of things, all but one of which belong to the same general class; the odd one is to be underlined, as in the following:

Silver; copper; glass; aluminum; gold.
81; 63; 49; 64; 16.

The second type of item involves statements which are either true or false and are to be marked accordingly:

Laminated armature cores are used because they retain magnetism better. True....False....

Resistance equivalent to a number of resistances in parallel is equal to the sum of the reciprocals of the separate resistances. True....False....

The third type involves problems of computation like the following:

What direct current of 110 volts will give the same horse power as a direct current of 5 amperes at 220 volts? Answer....

Given circuits of 4 and 6 ohms in parallel and in series with a circuit of 7.6 ohms, what current will be sent through by 120 volts? Answer....

Automobile Driver

One portion of this test concerns information about traffic rules and other matters conducive to safety.

1. If while driving you hear the gong of the fire department behind you, you should:
 -Drive faster in order to keep out of the way.
 -Drive more slowly to let the truck pass.
 -Drive immediately to the curb and stop.
 -Stop in the street as soon as you hear the gong.
2. The chief reason why you should avoid changing gears while crossing a railroad track is:
 -The tracks are rough and the bumping hard on the transmission.
 -You need all your attention to "stop, look, and listen."
 -Changing gears is liable to stall the engine.
 -You may get nervous and strip the differential.
3. Assume you are going to descend a steep slippery hill. Check three of the following things that you should do:
 -Leave the car in gear with the engine running.
 -Put the engine in reverse leaving the engine running.
 -Advance the spark lever.
 -Apply the foot brakes as necessary.
 -Put the engine in neutral.
 -Give the motor just enough gas to keep it running.

Another portion of the test involves recognition of dangerous situations. Pictures are shown on the blank and the subject is required to write what aspect of the scene is dangerous. The pictures include parking beside a hydrant or on a curve or double, passing a machine while ascending a hill and near the top, passing a stationary streetcar, traveling on the left side of a curve. The test also includes actual performance somewhat like the performance test for truck drivers (*infra*).

Performance Trade Test. As above suggested, performance trade tests are based on a quite different principle from most of the tests hitherto described. We have been dealing thus far, except for a few of the written tests, with the principle that if a man has worked at a trade for some time he will have picked up considerable information about it. The performance tests to

be described, however, deal with his actual ability to perform operations rather than with his information.

The procedure of selecting items on which the subject is to be scored is similar to that used in the information type of trade test. A preliminary set of tasks, tools, and material is gathered and items of score are devised. This tentative series is given to a few persons and then revised in the light of this preliminary try-out. When the final set of items is selected, it is possible to determine the critical scores in the same fashion as previously described. A little more ingenuity is often required, as, for instance, in selecting the aspects of the product to be measured and scored objectively. Care is also necessary to have supplies and equipment available and tools in good condition in order that everything will be standard. A few performance trade tests will be described.

Patternmaker

The applicant is provided with a standard set of tools and stock and with a blueprint. He is directed to "make a pattern for this cast steel bracket according to this drawing." The time is taken and the finished product is scored according to following standards. Various dimensions with their allowable margin of error are indicated on a photograph of the finished product. The applicant's product is measured to see how closely it conforms to specifications. If it falls outside any margin of error, a defect is scored against him. One dimension, for instance, must be between $5\frac{1}{2}$ " and $5\frac{3}{2}$ ", another between 4" and $4\frac{1}{16}$ ". A dimension outside these limits constitutes a defect. There are various other penalties, such as having the grain of one piece of wood run in the wrong direction or drilling the hole all the way through when it should go only part way; a total of 24 defects is possible in the finished product. A candidate is rated as a journeyman if his product has one of these defects and he completes the work in between 71 and 120 minutes. He is rated as a novice if his product does not consist of three or four blocks.

Interior Wireman

The applicant is provided with two joists and crosspieces fastened together to resemble a portion of a ceiling. He is also

given certain insulating tubes, knobs, wire, tape, and various tools. His instructions are as follows: "This is a part of a ceiling, joists, and crosspieces. Run two feed wires across and through both joists, using holes already drilled. From these main lines tap off leads in parallel and drop a lamp cord from this support. Use any material necessary, but do not use any more than you have to." The applicant is required to repeat his instructions in order to insure that he knows what is required, and is then left to his own devices. The finished product is scored according to a standard scheme. Certain aspects of the work are given one point credit if done in one way and no credit if done in another way. For instance, if the wires are drawn through the two outside holes 5" apart through both joists, the applicant receives a credit of one, while if they are drawn through holes less than 5" apart he is given no credit. If he leaves rubber tape or an open wire exposed, he receives no credit, but if friction tape entirely covers the rubber tape and all open wires are covered, he scores one point credit. He is given one point if the main lines are soldered tightly, but no credit if they are loosely soldered. In this way there are twelve possible items of score. An applicant is rated as a journeyman if he makes at least 9 points and finishes in less than 30 minutes. He is an apprentice if he makes between 2 and 8 points and takes more than 30 minutes. Less than 2 points indicates a novice.

Truck Drivers

The two foregoing instances are typical of the product-time test in which the subject takes a given test, the finished product from which can be scored. One illustration will be given of a process test in which the subject is required to put a truck through certain maneuvers. The examiner sits on the front seat beside the subject and scores him on certain aspects of driving during the test. After certain preliminary manipulations of levers and driving forward and backing in the open, the subject enters a course 9 feet wide marked off by stakes every 5 feet. The first portion of this course is in the shape of a letter S and the subject drives through at the speed he "thinks best." He is scored in this part of the test on the following errors: racing the engine when starting or shifting, starting abruptly, grinding the gears

when shifting, going through the S-shaped road in first speed, or knocking down a stake. At the end of this course he drives his hood between two posts that are rather close together. He is penalized if he knocks them down. He then has to back through a semicircular road without knocking down any stakes, an error being scored against him if he makes more than one direct backing in order to enter the half-circle or if he knocks down more than one stake. He next has to back the rear of the truck squarely up to the center of a small platform and further errors are scored if he hits the platform or approaches it at an angle. He then goes to another part of the course where he is required to turn around on a side hill. Possible errors include letting the truck roll downhill more than a foot, driving with the emergency brake on, backing more than once in order to turn around, or stalling the engine. After the subject has completed this course and all the errors have been noted, he is rated. An expert makes 3 errors or less, a journeyman from 4 to 9, an apprentice from 10 to 15, and a novice 16 or more.

GENERAL PRECAUTIONS

Reliability and Validity. Just as in a test of capacity, reliability and validity should be considered in a trade test. It is possible, after the final set of items has been selected, to divide it arbitrarily into two equal parts and determine whether the subjects make approximately the same score in the two parts. It is less satisfactory with this type of test to give it twice and compare initial and subsequent scores, because memory for items in the first test will influence the second score. Many subjects after the first test will look up or inquire about certain answers which they did not know and hence in a second test do much better. Those who have not done this will be at a disadvantage. Moreover, in many trade tests there are so few items that if one half is compared with the other half there is opportunity for considerable error because of the small number of items.

In some cases a trade test is constructed in two forms as is done with aptitude tests. It is then feasible to give both forms to the subjects and correlate them in order to determine reliability. The U.S. Employment Service has done this for 49 of

the 126 jobs for which they have standardized trade questions. Reliabilities are reported ranging from .86 to .91 [2, 44].

The validity of the trade test is largely revealed in the selection of questions and the calibration procedure above described. It is unsatisfactory to correlate scores with the criterion when the latter consists merely of the four degrees of trade ability, for the validity cannot be stated in quantitative form. With the graphic method of calibration, however, it can be seen whether each item and also the total of the items make possible a fair separation between the different degrees of trade ability. If this can be done with little overlapping, the test may be considered to have fairly high validity. It is, of course, also possible to give a trade test and then compare scores with success in the trade at a later time. If those who made high scores are doing successful work and some of them have perhaps been advanced to more responsible or supervisory positions while those with low scores are ineffective or have perhaps been dismissed, this gives a further check on the validity.

Recalibration in New Situation. Just as with the various tests and measures previously discussed, it is erroneous to assume that because a trade test worked in one particular situation, it will be of value in any remotely similar situation. It is desirable to recalibrate it in the place where it is going to be used. While any given trade has a good many fundamental facts and operations that will be involved wherever it is plied, there are many differences between organizations. A trade learned in one plant may differ in many essential respects from that same trade in another. For instance, the first plant may have archaic machinery while the second has modern equipment. The man who has worked and is skillful in the first may be at a loss in the second; and while such a man would make a high score in a trade test devised in the first plant, that same test would be unfair to workers in the second plant who deal with a different kind of machinery.

The importance of checking over the trade tests in a new situation is brought out by the fact that the U.S. Employment Service investigated many of the trade test questions devised for military use in 1918 and found about half of them unsatisfactory from the standpoint of validity. Some of this may be due to the

fact that materials and processes changed in twenty years. At any rate, it does point to the desirability of evaluating the tests in the new situation to be sure that they are really doing what they are supposed to do.

It may be desirable to start from the beginning, devise questions, revise them, and finally select and calibrate a set. Or it may be possible to take a set already developed elsewhere and see how valid it is in the new situation. In either instance a little research is necessary before the trade test can be made a valuable part of the employment program.

SUMMARY

Trade tests are designed to measure the ability possessed by a prospective employee at the time of application rather than any innate capacity that will enable him to achieve success after adequate training. They are not prognostic. They are needed in cases when it is unwise to take the applicant's word as to his trade experience and status. It is desirable to devise the test so that it can be administered by an examiner with no trade knowledge and so that it will yield an unequivocal and objective score that is quite independent of the judgment or knowledge of the person evaluating the results.

Trade tests are based on one of two general principles. It is possible to ascertain some information regarding a person's trade status by giving him a standard sample of work to do. It is also possible to obtain indirect indications by testing his information regarding details of the trade on the theory that an experienced worker will have incidentally picked up considerable information about his trade and will be familiar with tools, materials, and processes so that he can answer questions about them.

There are four common types of trade tests. In the oral type the questions are asked aloud and the subject's replies noted by the examiner. In the picture method the applicant is questioned regarding details in pictures of implements or machinery used in the trade. The content of the written test is similar to that of the oral test, but the form is such that the subject has merely to select the correct one from a group of alternative answers. The written test is usually adapted to group administration. In the performance test the applicant does some typical standardized

operation, perhaps on a small scale. This may be scored according to the process he uses, the finished product, the time consumed, or a combination of all of these. The oral and picture tests have the advantages accruing to other individual tests, namely, they minimize opportunities for misunderstanding and consequent erroneous results and make possible a certain amount of clinical observation. The picture test has the additional advantage of being more concrete and making greater appeal to the applicant, but it has the disadvantage that the picture may represent a different model of machine from that with which he is familiar and thus throw him completely off the track. The written test in a group form leads to great time-saving and the answers are unequivocal and can be easily and quickly scored by anyone. The performance test comes closer to the practical situation because it tests actual skill. It must usually, however, be given individually and this requires considerable outlay in the way of equipment and materials.

The method of developing and standardizing a trade test differs from that for innate capacity tests. Whereas in the latter there are a number of tests, each composed of many items, and the total number of items completed in a given time is compared with the criterion, in the trade test all the items are approximately the same sort and they are compared individually with the criterion to determine which are the most differential. The criterion that is most frequently used is a division of the subjects into novices, apprentices, journeymen, and experts, these terms being used in the conventional trade sense. Another classification consists of experts, apprentices and helpers, and workers in related trades.

It is necessary, by consulting technical sources and conferring with foremen, to devise a preliminary set of items of information or performance. These should be revised in order to clear up ambiguities. It is well to confer with foremen on this revision and also to give the items to a small group of workers in order to locate any misunderstandings. This preliminary set of revised items is then given to workers in the criterion groups, and the percentage of the members who answer a given item is determined for each group. If the percentage of the apprentices is higher than that of the novices, if the apprentices in turn are ex-

ceeded by the journeymen, and if the experts have the largest percentage of all, this particular item may be considered differential of trade ability. This determination can be facilitated by plotting a curve for these four percentages. A similar procedure is carried through for each question or item. It is then possible by inspecting the graphs to determine the most differential questions. Another possibility is to consider the statistical significance of the differences between the various percentages. The questions thus selected will be embodied in the final form of the trade test. It then remains to calibrate this final set of questions in order to set critical scores. This can be done graphically by plotting the total score of each individual, keeping the different trade classes in separate blocks of the chart and then drawing by inspection a line between the classes that will give the least possible overlapping.

It is desirable to investigate the reliability and validity of a trade test when this is possible. Half of the items may be correlated with the other half to determine reliability, although this is often not feasible because of the small number of items used. The validity is largely revealed in the calibration procedure, but it may be possible to compare scores with subsequent success in the work.

If a trade test has been developed in one situation, it is not safe to employ it in another similar one without further investigation. It often develops that methods of doing the work or the type of machinery used in one plant differ sufficiently from those in another so that a test devised in the former will be unsatisfactory in the latter. Whether or not this is true can be ascertained by repeating the calibration procedure to determine whether the critical scores hold in the new situation.

REFERENCES

1. Chapman, J. C. *Trade Tests*. New York, Holt, 1921, 435 pp.
2. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
3. Thompson, L. A., Jr., et al. *Interview Aids and Trade Test Questions for Employment Offices*. New York, Harper, 1936.

Chapter XV

JOB ANALYSIS

NATURE OF JOB ANALYSIS

Job analysis, as its name implies, comprises a consideration of the employer's contribution in the way of tools, material, pay, or general work situation, and of the workman's contribution in the way of skill, intellectual capacity, previous experience, or personal qualities. Job analysis is closely related to job specification or occupational description. The analysis is the means and the specification the end. After a detailed analysis has been conducted, the result is a series of specifications which can be used for various practical purposes. The analysis studies and ascertains the nature of the job, and the specifications reorganize this material into usable form.

PURPOSE

Job analysis is conducted for several purposes. The first of these is the improvement of methods of work. If it is desired to determine the most efficient way of doing a job, this may be facilitated by stating in standard quantitative form the different parts of the operation. One may wish to know, for instance, the time required to turn a taper, the distance a workman must reach for a wrench, or the time spent by a salesman in making out his reports and in other routine work. This information may make it possible to improve efficiency by eliminating wasted effort or by making technical adjustments.

A second purpose of job analysis is concerned with the health or safety of the employees. To this end study is made of various conditions such as ventilation or illumination or the proximity of dangerous machinery to various parts of the worker's body.

The aim of this type of analysis is to find where readjustments are necessary in the interest of safety and health.

A third purpose deals with more effective methods of training employees. The content of a worker's instruction can often be organized more scientifically. For instance, if the difficulties of the various operations are known, it may be feasible to teach the less difficult operations first. This plan is sometimes followed in training apprentices where the trade is divided into a number of subdivisions which are taught successively. Again, if the successful salesmen encourage their prospects to operate the adding machine themselves and ask plenty of "yes" questions, these facts may be passed along to the new men in the course of their training.

A fourth purpose of job analysis and one with which psychology is most concerned contributes to vocational adjustment. From this standpoint the work may be analyzed with reference to the duties, working conditions, pay, and relation to other kinds of work, and the worker may be analyzed with reference to his various qualifications, innate or acquired. This information makes the hiring process more effective because some of the things that are needed on the part of the worker are known; furthermore, he can be given such information about the job as may be necessary to "sell" it to him or at least guard against ultimate dissatisfaction on his part because of initial misunderstanding. The results of the job analysis also may contribute to vocational counseling. Even without actual tests it will tell approximately what the requirements of the job are in the way of educational background, experience, strength, and the like. The counselor is in a better position to advise individuals. In fact, the U.S. Employment Service has carried out an extensive project for analyzing a large number of jobs with a view to more effective counseling and placement in the various centers it operates throughout the country [6].

NEED

The need for job analysis is quite apparent. Many occupational terms are ambiguous. For example, an applicant who indicated on his personnel blank that he was a pipe cutter was uncritically assigned to a job of laying sewer pipes. It developed subse-

quently that he had been a carver of Meerschaum pipes. Similar ambiguities are liable to occur in almost any industry. If a request is made for a machinist, men may be available who are good at lathe work but poor at bench work, or who can operate a drill press but are unable to do other kinds of machine work efficiently. If a clerical worker is desired, it is necessary to specify more than this general term indicates because qualifications are quite different for transcribing clerks such as timekeepers, bill clerks, bookkeepers, stenographic clerks who do shorthand and typing or secretarial work, filing clerks, clerks who meet the public at a cashier's window, or machine-operating clerks whose work is confined largely to computing machines. Hence it is obviously necessary to specify in somewhat more detail the actual nature of the job and the actual qualifications desired for that job.

THE ROLE OF PSYCHOLOGY IN JOB ANALYSIS

Use of Psychological Categories. Job analysis, to be sure, involves many things besides psychology. Much of the information deals with various items of industrial practice, but some of it also runs into psychological categories, especially when the necessary qualifications of workers are described. Mention is often made of an operative's innate capacity such as intelligence or attention. Such characteristics, we have seen earlier, may be approached more objectively, if desired, by mental tests. Account may also be taken of his acquired proficiency in various lines and this may be approached by the trade test technique already described. Again, the qualifications may include certain personality traits and these may be evaluated by the rating scale procedure. In other words, the description of the worker will often extend into psychological categories and may sometimes actually comprise the results of technical procedures such as have been discussed earlier in this book. A final job specification may frequently include critical scores on certain tests or rating scales.

Psychological Background for Job Analyst. Psychological training will probably help the job analyst. The psychologist usually learns to observe people somewhat more closely than does the ordinary individual. In a clinic, for instance, considerable stress is attached to the involuntary movements an individual makes,

the way he goes at a task, and the fleeting evidences of emotional abnormality. A person with a psychological or clinical background probably will observe whether the worker is performing his task automatically or with apparent conscious effort, whether he takes advantage of the rhythm of the operation, whether his eye necessarily follows his hand in making certain adjustments, whether a salesman dominates the prospect in the sales interview. Psychological training further helps in directing the analyst's attention to what the man does as well as to what the machine does. The casual observer is perhaps more inclined to watch the machine, whereas the psychologist will pay a considerable amount of attention to the workman. Moreover, this type of training makes one specially conscious of the necessity for concrete and specific descriptions. The psychologist is well aware of the limitations in terminology when dealing with human traits. Finally, while the technique of weighting different variables, such as items of personal history or test scores, in order to predict validly some other variable, such as occupational efficiency, is not unique with psychology, nevertheless the psychologist is usually familiar with this technique and hence has a rather good background for research work. In the following discussion a brief account will be given of current methods of job analysis followed by a consideration of its primarily psychological aspects.

METHOD OF SECURING DATA

Interview Procedure. A widely used method for securing job analysis data is to interview persons supervising the job under investigation or even the workers themselves. Earlier methods sometimes resorted to a questionnaire directed to either of these groups, asking them to state the nature of the job or the qualifications they thought necessary for it, or giving them a list of duties and qualifications from which to make their selection. This procedure was none too satisfactory because the persons filling out the blank often were not aware of the importance of scientific exactness and were inclined to use general and undefined terms. It is much more satisfactory, therefore, to have a trained interviewer secure this information by personal contact. When he is face to face with the supervisor or workman he can adapt his

procedure to the circumstances. If the interviewee is indefinite on a particular point he can be questioned further while that point is still under consideration. If a particular lead is given which appears quite significant to the interviewer, he can secure more information on the spot, whereas it might be difficult subsequently to check back with a second questionnaire dealing with that specific point.

Observation of Workers. The analyst may likewise secure valuable information simply from observing workers on the job. This is the same procedure as that followed when analyzing the job with a view to selecting tests (Cf. Chapter VIII). A skilled analyst will secure much information in this way. His observation will be appreciably facilitated if he is provided with some form specifying the special qualifications for which he is to watch. Some of these forms will be discussed below.

Analyst's Personal Qualities. The job analyst needs certain qualifications in order to do his work successfully, and various lists of them have been suggested [2]. For instance, he should have a rather high degree of intelligence, and ability to analyze the situation, to be alert for leads, and to discriminate the important from the unimportant. There is less certainty as to whether he requires technical training in the job. He must without question be sufficiently familiar with the work to understand its terminology. It would be absurd for an interviewer to approach a man and be unable to talk to him in his own language. If the worker uses terms which are familiar to himself but the interviewer repeatedly has to have them explained, it puts the latter in the position of not knowing his business and is conducive to lack of confidence. However, it is doubtful if the interviewer needs the familiarity with the occupation that comes from personal experience; it is possible for him to be even too familiar. There is a danger in the latter case of his going into minutiae that are insignificant from the practical standpoint.

In addition to intelligence and knowledge of technical terminology the interviewer should have various personal qualities. He must have patience because his work will often involve considerable delay. There may be unavoidable interruptions when he is about to interview an executive. Moreover, he needs tact for it

is often difficult to get a man to talk about his job. Some persons are more or less jealous in this respect and are apt to be reticent. If, however, the interviewer tactfully expresses interest in the man's work, he will probably be able to extract the desired information. He should also be rather persistent and firm in his manner because it is often necessary to keep the worker on the track. He may frequently have to say, "That is all very interesting, but now what about this?" The interviewer further needs to be able to inspire confidence and cooperation so that the men will be interested in helping him in every way that they can—in other words, he must be a good salesman. When a worker hails the man who interviewed him a few days previously and tells him that he has thought of one or two other aspects of his work that he forgot to mention earlier, it is obvious that this man has confidence in the interviewer and is anxious to cooperate. Sometimes one of them brings in notes he has jotted down as a basis for further discussion.

The job analyst needs further to be a good observer, as he may be called upon to watch persons at work and to decide whether they need manual dexterity, memory for details, or emotional stability. Good laboratory training in science and particularly in psychology should be helpful in developing this characteristic in the analyst.

Analyst's Training. The analyst must have some preliminary training before much value can be attached to his results, whether he is to conduct interviews or observe operations. The amount varies in different situations. For a survey at one of the government air service experiment stations one day's intensive training was given the interviewers, whereas one month's instruction was given as a preliminary to an analysis of secretarial work. This training may involve preparing questions and revising the wording of the questions so as to bring out the desired information. Trial interviews are often valuable; here the person conducts a few interviews, not with the intention of obtaining valuable information, but for the purpose of getting experience himself. The instructor can go over the results of these interviews with him and show him his mistakes and the good features. While preliminary training of this sort is usually given, it is not to be assumed that after it the interviewer can work entirely inde-

pendently. In a large organization where there are several interviewers they should confer frequently about their work with those to whom they are directly responsible.

In training analysts to observe workers the program of the U.S. Employment Service is typical. It begins with a review of the objectives to be obtained by the rating form so as to convince each individual of the importance of securing accurate data. The analysts are then given the standard work sheet (*infra*) listing the characteristics of the workers that are to be observed and rated. The definitions of these characteristics are read, discussed, and illustrated. Thereupon the analysts go out and analyze a fairly complicated job, using the technique as best they can. The results are discussed with them individually and compared with those of other members of the group. Points on which the different individuals disagree markedly are given special discussion. Incidentally, this procedure of securing several ratings of the same job is carried out on later occasions to insure that all the analysts are working in a similar manner.

Whom to Interview or Observe. If an organization is confronted with the problem of analyzing certain occupations, the next point to consider is what persons are to be interviewed or observed. If the technique is to consist of observation, efforts obviously must be confined to workers. If, however, interviews are to be held, there are two possibilities—the workers and the men who supervise their work. It might seem offhand that the superiors ought to know in great detail just what the men are doing and hence would be the most desirable men to interview. As a matter of fact, there are often minor aspects of the day's routine that do not reach the supervisor at all. For instance, a superintendent of a pressroom would consider that his foremen were essentially engaged in carrying out his orders and getting the work out on time; he might entirely overlook the fact that they also had to see that the presses were washed and oiled before they left each night. This operation is an important part of their job, but in an actual interview it did not occur to the superintendent. On the other hand, the worker may not give all the information desired. It is hard for a person to take a detached point of view toward his work and describe all its details. This is particularly the case if he has been at a job for a long time, for

many operations become relatively automatic so that he performs them with very little attention. Consequently, as he thinks back over his work with a view to analysis, he is somewhat less apt to recall the aspects which do not occupy much attention during the day's work. Hence it would seem desirable to secure information from both the workers and their superiors, trusting that the details omitted by one will be supplied by the other.

It is further desirable to have a typical sampling of individuals for the interview. In some instances it is, of course, possible to include everybody in the concern who is working at a given job as well as all the supervisors. If this is not feasible and a sampling is to be taken, it is well to insure that the sampling is typical and does not represent a special aspect of the work. In a study of secretarial workers in which persons in a great many establishments were interviewed, effort was made to sample those in four different lines of work—secretaries in general business capacities, secretaries in government positions, secretaries in institutions, and secretaries to professional men [2]. When the samples were selected in this way there was less danger that the analysis would reflect the peculiar features of one particular kind of secretarial work. Identical principles are involved in the sampling of workers to be observed.

No definite rule can be laid down as to the number of people who should be interviewed or observed. After the procedure has reached a certain point, it will become obvious that the last few individuals have contributed nothing in addition to what has been contributed by earlier ones. Consequently, further investigation will probably be of little value because it will yield little additional information.

Work Sheets. A printed form or work sheet for securing the preliminary job analysis data is standard practice. It is inadvisable to attempt to write up an occupational description on the spot. It is better to gather the information in a systematic fashion and then organize it at leisure. A number of typical work sheets or forms will be described.

The U.S. Employment Service has a "Worker's Characteristics" form. This is a list of about 50 characteristics, with provision for a symbol rating of each one. Some of the characteristics are as follows: (1) working rapidly for long periods, (2) strength, (3)

dexterity, (4) coordination, (5) estimations of various sorts such as size, quantity, or speed, (6) special senses, (7) memory, (8) social factors such as tact, (9) temperament such as emotional stability, (10) miscellaneous factors such as ability to make decisions, oral expression. Subdivisions of these characteristics are frequent.

The analyst has this sheet with him while observing the worker and makes his judgments on the points indicated. With this particular sheet he rates each characteristic as A, B, or C. A rating of A means that an unusually large amount of the trait is demanded by the job, such as would be possessed by not more than 2 per cent of the general population. A rating of B means a distinctly above average amount such as would be possessed by the next 28 per cent of the population. The C rating means that the amount of the characteristic is less than that possessed by the highest 30 per cent of the general population. In making his estimate the analyst thus has to compare the workers with persons in general. He is also provided with a manual which defines the characteristics and usually illustrates them in a specific job. For example, with the item "sense of taste," the definition stresses ability to distinguish differences in quality and intensity. It is illustrated by the roasting foreman's job in chocolate manufacture or confectionery manufacture. The elements in the job that necessitate taste are described—for instance, chewing a sample of cocoa beans to determine whether they are properly roasted.

Another form for securing data about the job is somewhat similar to the foregoing but has some additional features [1]. The first item deals with "mental effort," including general and special education, monotony, instructing other people, preparing records, intelligence, patience. The next item involves skill; the kind is to be specified, likewise any desirable prior experience. The third item deals with physical aspects—age limits, height, weight, sex, sensory factors such as vision, the amount and nature of the physical effort required, and the amount of fatigue involved. The fourth division concerns responsibility carried by the individual such as for equipment, tools, materials, property, or the work of others. The final classification deals with working conditions, with subitems covering place, type of surroundings, atmosphere, illumination, and hazards.

Another form that may be mentioned comprises a detailed check list on certain points [3]. For instance, under personal conditions there is: work with others, close to others, direct others, plan work. Under working conditions is the item machinery with subclasses: floor, hand, electric, bench, foot, care of. Under tools there is: light, heavy, furnished by the worker. Materials include: light, heavy, fine, coarse, perishable, care of. Motion includes: simple, complex, fast, slow, large, automatic, varied, repetitive, rhythmic, small. This particular form lists a large number of items which, however, are merely to be checked rather than rated on any scale as in the case of the first form described above.

One of the more elaborate of the earlier work sheets may be described [4, 141]. The first page contains space for the description of the work—the duties, responsibilities, tools and equipment, and working conditions. In describing the duties emphasis is placed not on mere detail, but on a statement of the functions of the job. “Responsibilities” includes such things as custody of money or property and insuring the safety of other employees. Under “tools and equipment” are to be mentioned not such tools as hammers or shovels that anyone can handle without much special instruction, but rather things involving special skill and training such as typewriters or welding machinery. To facilitate the evaluation of working conditions a code is appended as follows:

- A. Imminent risk of life; e.g., experimental parachute jumper.
- B. Dangerous; e.g., propeller tester.
- C. Hazardous; e.g., aviation mechanic, ground man.
- D. Unhealthy or extremely unpleasant; e.g., doper, propeller tester.
- E. Factory or shop.
- F. Office.

The notation of one of these code letters on the sheet is all that is necessary.

The next page of the blank deals with a set of minimum requirements on the part of the worker. Physical qualities are coded in somewhat similar fashion to the preceding, as follows:

- A. Superlative; e.g., great strength (continuous heavy lifting), exceptional eyesight (draftsman, instrument maker).
- B. Superior; e.g., unusual strength (occasional heavy lifting); good eyesight (machinist).
- C. Better than average; e.g., better than average strength (carpenter, plumber); better than average eyesight (typist, fabric worker).
- D. Below average—average strength not needed (watchman, messenger, engineer); average vision not needed (doper, dry kiln operator, fire fighter).
- E. Slight—little strength required (office worker, draftsman); poor vision acceptable (janitor, laborer).

Another item concerns education and has a space for entry somewhat similar to the graphic rating scale:

POST-GRADUATE																								
WORK					COLLEGE					HIGH SCHOOL					COMMON SCHOOL									
V	IV	III	II	I	G	IV	III	II	I	G	IV	III	II	I	G	8	7	6	5	4	3			

The number indicates the grade of common school, high school, or college the individual finished. In a similar way data regarding requirements for special training or experience may be recorded:

Special training—	V	IV	III	II	18	12	6	3	1	none
Experience —	V	IV	III	II	18	12	6	3	1	none

The arabic numerals indicate months and the roman numerals years. With reference to technical skill the sheet provides a line comprising the four usual trade classifications:

Expert Journeyman Apprentice Novice

The presence of this item on the blank suggests, of course, the desirability in some instances of setting a critical trade test score. If the results of this job analysis are to be used in employing persons where technical skill is desirable, it will be more satisfactory, as has been shown previously, actually to give a man a trade test and determine on that basis whether he has the requisite trade ability than to take his word for it. The job analysis would then state the amount of technical skill necessary, and the trade test would determine whether the applicant had the requisite skill.

Further items on the blank deal with personal qualities and are arranged like a graphic rating scale:

Judgment	Unfailing; errors cause personal danger		Good; errors cause money loss		Average; errors cause confusion		None
Creative ability	Highest; inventiveness		High; originality		Average; initiative		None
Number supervised	500	100	25	10	2	None	

Each of these "requirements" has also a blank space labeled "reason," in which the analyst must justify the entry he has made. For example, on the work sheet for automobile mechanics the entry "keen hearing" was justified by the statement that this was necessary in order to diagnose motor trouble; common school education was required in order to make out time slips and read written directions; a year's previous training in a garage or repair shop was requisite in order to shorten the learning period, and good judgment was listed because it was required in "shooting trouble." This procedure of making the analyst justify each entry insures that the item listed represents a real requirement and not an imaginary one. It puts the analyst and, in the case of interviews, the worker or executive to the necessity of really considering the value of certain items. It also clarifies the qualification itself by showing a concrete way in which it is to function.

Another page of the blank is similar to the one for minimum requirements, but deals with further requirements that are desirable but not absolutely essential. It comprises the same set of items with spaces for writing the answers and also justifying them. The interviewer can then list the qualifications according to whether they are essential or simply desirable.

Many of the items on such work sheets are not psychological in character, but there are manifestly certain aspects in which psychology is or might well be involved. The conventional rating scale procedure is suggested by the consideration of various character traits. The question of trade qualifications immediately points to the technique of trade tests. In certain types of work additional items regarding intelligence might prove desirable. It might be possible to analyze a job with reference to whether it required a high degree of attention, a certain amount of

memory, the ability to make quick decisions, or other special capacities. A person with psychological training might frequently find items of this sort which could well be included in the analysis of the job.

OCCUPATIONAL DESCRIPTION

Form. After a considerable number of persons have been interviewed or observed and a work sheet filled out for each, the analyst can note the statements regarding duties and qualifications on which there is substantial agreement. He is then in a position to write up the results of his interviews in the form of a final occupational description or job specification. While the form of this description may vary with the circumstances and the preferences of those most concerned, it is a rather established practice to put the description in simple declarative sentences.

Examples. A few typical occupational descriptions will be cited.

OCCUPATIONAL DESCRIPTION FOR AUTOMOBILE MECHANIC¹

Duties	The automobiles and trucks used by this company are kept in condition in the Garage Branch of the Maintenance Section. Under direction the automobile mechanic overhauls, repairs, and operates such standard machines as the Dodge and Cadillac touring cars and Mack, Standard B, and G.M.C. motor trucks. He tests, overhauls, and repairs motors, generators, and ignition units. He does acetylene welding, and uses tools such as lathe, reamer, and valve-reader.	
Hours	7:45 A.M. to 11:30 A.M. 11:30 A.M. to 12:15 P.M. Lunch 12:15 P.M. to 4:30 P.M. 7:45 A.M. to 11:45 A.M. Saturday.	} Monday to Friday.
Minimum qualifications	The automobile mechanic must have graduated from common school and in addition he must have had three years' practical experience in a garage or automotive machine shop as repairman. In lieu of one year of practical experience, six months' special training in automobile repairing or one year as machinist apprentice will be accepted. Man 18 to 50 years of age.	

¹ After Scott and Clothier.

Additional qualifications desired The automobile mechanic should be physically strong, capable of occasional heavy lifting. He should have good eyesight in order to do close work and make fine adjustments, although glasses are permitted. Keen hearing is also desired in order to enable him to test motors by sound.

Accuracy is important in this work as errors may cause delay and impair work.

Working conditions Garage with concrete floor. The worker is on his feet about half the time. Much of his time is in a crouching or prone position incident to repairs underneath cars. The automobile mechanic is outdoors part of the time, especially when testing machines on the road.

Principal lines of promotion From: Truck driver, chauffeur, mechanic's helper.
To: Garage superintendent, engine mechanic.

It is obvious that this description embodies the information discussed earlier in connection with the interview. It begins with a description of duties, stating exactly what the man does. It also gives the hours which he works. The next section states the minimum qualifications and covers the various topics of education, experience, and the like that have been discussed before. There are also additional qualifications which are desired but not absolutely necessary. Further information is presented regarding working conditions and also the principal lines of promotion. This latter gives a notion as to the most profitable positions from which to recruit personnel for the job in question and also the positions to which a man may be promoted after having had adequate experience.

In the original occupational description sheet just presented there is also a series of boxes at the top for quick reference. These boxes deal with such items as education, experience, judgment, accuracy, supervision, physical qualities, and working conditions. In each box is entered a single letter which refers in code to different degrees of the particular qualification or item. The codes for physical qualities and working conditions have already been given (p. 517). The following are the notations for recording the remaining items in code:

EDUCATION

- A. Graduation from college.
- B. Graduation from high school.

- C. Two years' high school.
- D. Graduation from common school.
- E. Six years' common school.
- F. None.

EXPERIENCE

- A. Ten years.
- B. Five years.
- C. Three years.
- D. Two years.
- E. One year.
- F. None.

JUDGMENT

- A. Errors may cause loss of life.
- B. Errors may cause personal injury.
- C. Errors may cause money loss.
- D. Errors may cause confusion—interdepartmental.
- E. Errors may cause inconvenience—intrdepartmental.
- F. None.

ACCURACY

- A. Errors may cause loss of life.
- B. Errors may cause personal injury.
- C. Errors may cause money loss.
- D. Errors may cause confusion—interdepartmental.
- E. Errors may cause inconvenience—intrdepartmental.
- F. None.

SUPERVISION

- A. Supervising 100.
- B. Supervising 50.
- C. Supervising 25.
- D. Supervising 10.
- E. Supervising 5.
- F. None.

OCCUPATIONAL DESCRIPTION FOR DESIGNER IN STRUCTURAL STEEL²

Duties and Responsibilities The designer in structural steel designs the steelwork necessary for *cooling towers* and *cooling bridges*; the steel framework for *substations* and *generating stations* and miscellaneous structures such as stairways and platforms. All

² After Snow.

of this work is constructed either by contract or by the company's building construction department.

In designing steelwork the designer should be familiar with:

- a. The loads to which the structures will be subjected.
- b. Structural steel handbooks which give tables of the size of structural members, such as *eye beams*, *H columns*, *channels*, *angles*, and *girder beams*, which he will use.
- c. The standard methods of making connections between *beams* and *columns*, *beams with beams*, etc.
- d. The necessary *struts* and *braces* and the methods of connecting these with columns and beams.

The designer in structural steel also designs the steel framework for additions to be made to existing buildings. Before making these designs he takes field measurements at points where the new work is to be added, noting carefully whether reinforcements will be necessary in any existing construction and such things as connections of floors and walls and special foundations for heavy machinery.

Where steel stairways, platforms, or ladders are to be built inside of substations or power houses, the designer takes field measurements at the location, allowing for clearance between new and existing work.

At times the designer designs *steel smoke breechings* for boiler rooms of power houses. In making such designs he should make allowances for expansion of the breeching due to wear, seeing that the proper clearances are allowed between these structures and existing steelwork. He should be familiar with *expansion joints* necessary for *smoke breechings*.

He makes preliminary layouts of all designs; he should be familiar with standard drawing practices, structural steel designs, and standard drawing instruments. He must know how to operate a Universal drafting machine and use all miscellaneous materials used by draftsmen such as drawing paper, tracing cloth, different grades of pencils, and drawing inks.

Personal
qualities
desired

A man of 25 to 35 years of age is desired .

Initiative above the average is essential, as he must in most cases use his own judgment in working out the best methods of design.

Accuracy in determining the necessary size and kind of structural member is of prime importance, as a considerable saving of money is effected if these members are of the exact weight necessary to support a definite load plus an additional load for a factor of safety.

Neatness in making drawings that can be understood by others is an essential quality.

Working conditions The work is permanent and is highly technical and mechanical in nature. Working at a drawing board making drawings may cause some eyestrain. The drafting room where the layouts and designs are made is well lighted, ventilated, and arranged.

There is some outdoor work attached to this position as when taking field measurements.

Education and experience desired A four-year technical education or its equivalent is desired. One that is specialized on the theory of structural steel design is preferred.

Two years' experience designing steel buildings for power plants and substations is desired.

Opportunities for advancement There is at present no direct line of promotion from this position.

There is opportunity for men of this type to secure positions with high responsibilities with structural steel corporations.

Sources of supply Draftsmen of structural steel corporations are external sources of supply.

Draftsmen are an internal source of supply.

This description is in form substantially like the preceding, giving information regarding duties and personal qualities, working conditions, education, and experience, and lines of promotion or positions from which to select people for promotion to this position.

SUMMARY OF ANALYSIS OF WATER GAS MAKER'S JOB³

General Description

(Based on Activities and Products Made)

General Description of Duties: Produces blue gas. Supplies coke or coal to generators and passes air and steam through coke to make

³ After Williams [8].

water gas to heat coke ovens where oven gas and coke are produced and also to mix with oven gas for city use.

Produces Coke. Supplies coke or anthracite coal to water gas generator; admits steam and air to make gas and pumps gas to heat foundation ovens for making coke.

Produces City Gas. Pumps gas to holder to mix with oven gas in order to produce required quality and quantity of gas for consumers.

Summary of Duties

Charges Generator. Turns screw lock and opens lid of generator, pushes lid away (on rails), pulls on larry car by hand to roll over hole, pulls gear check loose with fingers, moves larry back and forth over hole by hand crank (handle on wheel) and turns wheel, on side, to close bottom. Rolls larry away and rolls lid (above) over hole. Screws lid on tight by hand.

Starts Automatic Operation. Pulls lever out on control board to start air blast. Looks through peep sight until color of fire shows cherry red. Pushes levers on control board to admit steam after set has been shut down for any time. Pushes small handle on side to "split-run" position after first "up-run."

Charges Relief Unit. Looks at fire in relief generator and charges as necessary.

Inspects Seal Pot. Turns valves on seal pot at rear of boilers, waits for liquor to drain, inserts short rod to determine sludge level, closes valve.

Informs Foreman. Reports unusual conditions to foreman.

Records Operation. Records major operations in log book to provide written record.

Inspects Water Level. Looks at level of water in water column on boiler.

Supervises Fire Cleaning. Looks over generator carefully after cleaning by fire crew and passes or rejects job depending upon amount of clinker left.

Physical Characteristics or Factors of Operator

Physical Alertness. Walks to various positions on one level (operating floor) and maintains constant alertness during entire shift.

Physical Strength. Handles larry car, moves lid and assists fire crew. Handles simple levers and valves, some strength required.

Physical Endurance. Works entire shift of operations listed. Some (infrequent) climbing of stairs or ladders required, work in standing or walking position with some rest periods.

Physical Health. Freedom from:

1. Impairment of vision or hearing
2. Impairment of legs
3. Heart disease

Experience

Minimum Experience. Helper—one year. Prior to Employment in Plant: General experience as steam engineer or fireman on railroad, in power plant and similar jobs.

(*Minimum*) *Learning Period to Perform Duties Without Direct Supervision on Job.* Working with experienced Water Gas Maker.

Relation to Other Jobs. Promotion to general foreman. Rotation to producer gas operator, heater. Promotion from turbine operator.

Abilities

Eyesight. Read $\frac{1}{2}$ " numbers on gauges at 3'. Read approximate position of 12" hand on 14" face at 75'. Read 15" clock at 50'. Distinguish between light blue, cherry red and purple.

Touch. Discriminate by touch and feel for change of 50° Fahrenheit on metal casing.

Accurate Movements. Use wrench to connect pipe.

Figuring (Arithmetic). Differentiate numbers 1-100. Weight relation—pounds and tons. Read multiple record chart in temperature and pressure units. Add, subtract 4 digits.

Oral Memory. Repeat meaning six simple sentences after one hour span (context).

Visual Memory. Read six gauges and repeat reading without error after 5 minutes.

Observation. Notice brick (6" x 2") 2" out of line at 10' distance.

Multiple Operation. Watch gauge and turn valve simultaneously. React to auditory signal (bell) while working on job. Watch movements of levers at distance of 50' and position of indicator on gauge.

Hearing. Distinguish difference in pitch of turbine. Notice when usual starting noise of turbine fails to occur every 3 minutes.

Understanding Instructions. (1) Oral—Give and understand simple verbal orders. (2) Write and read simple description as contained in log book and manual of instructions.

Work Situation

Place. Inside water gas building on second floor.

Materials. Coke and coal stored in bins and supplied to generators by manually controlled larry car. Handled almost entirely by helper.

Exposure to Weather. Entire shift spent on operating floor of water

gas building with complete protection from weather.

Moisture. Water in small amounts on ground floor.

Temperature. Very hot in summer.

Safety Precautions

Precautions with Regard to Hazards

Process. Possibility of explosion at start of operation due to having down run or back run instead of up run or failure to purge.

Personal. Gas or combustion burns: Possible injury from back draft while removing lid from generator if not careful. Falling objects: Possible injury from falling coke.

The foregoing are illustrations of rather detailed job specifications prepared along the lines described. Mention may be made again of the *Dictionary of Occupational Titles* prepared by the U.S. Employment Service [7]. This provides brief job specifications for 20,000 occupations. It was designed for use by vocational counselors and placement offices, but the job specifications, though brief, may be useful for some industrial purposes. A few of them may be cited by way of illustration.

Shoe-fitter, custom made (boot and shoe). Operates various types of sewing machines to sew complete shoe uppers and is responsible for their correct fitting; skives all parts with a thin-bladed knife; cements linings together with rubber cement; closes back seams of shoe uppers on sewing machine; fits vamp on back quarter of shoe, using a wooden block and pasting parts together with rubber cement; sews lining to uppers on a sewing machine; trims all parts of shoe uppers with pair of scissors. May perforate shoe uppers on perforating machine.

Tire builder, core. Builds large passenger car and truck pneumatic tires by hand on a mechanically turned core; (1) applies bands on core; starts band on circumference of core by hand; presses control pedal to revolve core; guides band to center it on the core by pulling a roller against the inner side of band; shapes edges of band to core by working a stitcher (small roller) and a spade (flat-ended hand tool) over the top and outward toward edges of band; applies as many as six bands to large tires; (2) builds in beads on edges of tire; lays bead in position along edge of band (usually applied over second band); folds edges of band around bead using stitcher and spade, as core revolves, taking care to make it smooth; (3) applies cushion band, tread rubber and side wall rubber using stitcher, spade and

hand. Pulls built tire from core after turning hand wheel that collapses core, and lays tire aside ready for curing.

PSYCHOLOGICAL POSSIBILITIES IN JOB ANALYSIS FOR EMPLOYMENT PURPOSES

From the above discussion of job analysis it is evident that some of the points are distinctly psychological in character. In so far as the job specification deals with the worker, it is bound to include mental factors and psychological terminology. This is related to the interests of the psychologist in two ways. As hinted in Chapter VIII, the results of job analysis may be of value to a psychologist who is initiating a project for developing mental tests for a particular occupation. In such a case he must determine what mental characteristics are necessary for the occupation with a view to devising tests for those characteristics. If a careful job analysis has been conducted, the psychologist may find it a valuable starting point for his own analysis for developing tests. If the occupational description mentions keen hearing, good attention, powers of observation, or the necessity of making motions quickly, this points to rather obvious psychological test possibilities. The experimenter will doubtless supplement this type of information with further observation of his own, but it often calls his attention to aspects of the job that he might otherwise have overlooked and affords him a good beginning for his work.

On the other hand, the psychologist has something to contribute to the job analysis program. Many of the principles discussed earlier in this book might well be considered here as a supplement to the method. If the job specification is to be the final instrument used for hiring workers, it might theoretically embody a number of these principles. The remainder of the chapter will point out a few which might fit into a comprehensive job analysis program.

Statistical Validation of Miscellaneous Factors. In the first place, it may often be desirable to evaluate statistically certain miscellaneous items of personal history such as are brought out in the analysis. For instance, height and weight are sometimes noted by analysts as desirable for a given kind of work, and if they attempt to justify such items on the work sheet they will

state that the work is "heavy." Or if the analyst finds that an eighth-grade education is necessary, he justifies it on the grounds that the worker must read time slips. If he says that a married worker is preferable, he substantiates this judgment by the fact that such a worker will be more stable.

It is statistically possible to find out whether a certain height or weight is necessary for the job, whether an eighth-grade education actually is the necessary minimum, and whether married workers are more stable. The procedure discussed in Chapter XIII on miscellaneous determinants of vocational aptitude is directly applicable here. It is necessary merely to obtain groups of workers of a given type, some of whom are reasonably successful and others unsuccessful, and tabulate them with reference to such items as height, weight, age, or marital status, to see to what extent these items differentiate the successful from the unsuccessful group. While the judgment of the analyst may be sound when dealing with matters that are fairly obvious, there is no real guarantee that his information is always well founded. Some of the persons whom he interviews may have made hasty generalizations and passed them on to their colleagues, so that there will be unanimity in a statement that is actually erroneous. The technique of statistical validation will insure against any such error.

Rating Scales. In the second place, the technique of rating scales would seem rather generally applicable to the various personality factors that are sometimes encountered in job specifications. Some of the work sheets described above embody as much of this technique as is applicable. Statements regarding judgment or creative ability are recorded by checking on a line with descriptive phrases beneath it. The only suggestion to be made at this point is a somewhat wider extension of this technique to cover other traits that might be of possible significance in many types of work. Such traits as tact, leadership, cooperativeness, and many others discussed in Chapter XII would seem applicable. The average location of the check marks on the work sheets would indicate the degree of the trait that was requisite. In some cases the ratings might be combined quantitatively into a total rating as described previously. In many instances, however, they could be recorded in code similar to that mentioned in

the present chapter. It would thus be possible to determine something analogous to critical scores in many specific traits and embody them in the final job specification. In using these facts for promotion or transfer within the organization, those being considered for such change could be rated systematically by their superiors. In hiring employees from outside, the use of a rating scale during the employment interview might be adopted.

Trade Tests. In the third place, the possibility of trade tests has already been suggested. Whenever the job specification asks for technical skill and mentions apprentice or journeyman or expert, this immediately raises the question of how this trade status is to be determined. To be sure, a man's union card will often give an approximate notion of it. There is no guarantee, however, that it will always be reliable, and it is more dubious still to take a man's own word in the matter. The technique of trade tests has reached the point where it would be applicable to almost any type of operation that requires specific trade skill. Consequently, if such tests were developed for the job in question the occupational description might well include a critical trade test score.

Intelligence Tests. In the fourth place, intelligence tests might well be an item in some job specifications. We have previously seen that in certain work, such as clerical, there is a definite correlation between intelligence and occupational efficiency. We have also noted that some occupations in the hierarchy require a certain general level of intelligence and that persons too low or too high are unsuited for those types of work. In such cases it is customary to establish a critical score in intelligence as a basis for hiring. It would thus seem logical, in an organization where intelligence tests had been standardized, to embody in the job specification a critical score in the intelligence test.

Special Capacity Tests. Finally, the tests of special capacity which were described at considerable length in Chapters VIII and IX might play a role in this procedure. In any occupation for which such tests have been worked out, critical scores either on separate tests or on the weighted sum of the tests might be introduced as one important item in the job specification. The aim of job analysis and specification is, of course, to present to the applicant all the necessary information about the proposed

job and its possibilities, and to obtain all the necessary information about him with a view to occupational prognosis. While a given organization must be governed considerably by the extent to which it can invest in the employment program, and while the validity of psychological methods depends considerably on the local situation, there are doubtless a great many instances in which it would be desirable to develop a rather complete and extensive job specification. This comprehensive specification would include items similar to those in the various specifications cited above, and might also comprise critical scores in rating scales, trade tests, and various other tests of innate capacity which have been statistically studied with reference to the job in question.

“Families” of Occupations. An additional aspect of job analysis procedure has been explored by the U.S. Employment Service. After making the analyses of the jobs in terms of 50-odd characteristics, they investigated the relations between the different jobs with regard to these characteristics, the purpose being to discover certain types of work that are fundamentally related—“families” of occupations. Information of this sort will be valuable in cases where in lieu of giving actual tests it is desired to employ a person from a related job. If no experienced tool-makers are available, for example, are there other jobs that so closely resemble this that men with experience in them will quickly acquire proficiency in tool-making? In other words, what are the best sources from which to recruit workers for a given occupation if it is desired to lose as little time as possible in breaking them in to the new job? Much of the analysis is made by means of a punched card system that uses the rating procedure for the various qualities mentioned above (p. 516). It is possible to develop certain families of occupations that have a good deal in common. For instance, the following were included in one family that required B grade of strength and dexterity, a training period of one week, but no minimum formal education or special knowledge: circular-saw operator, granite polisher, gritter, blossom maker—automatic, paper cutter, leveling-machine operator. Another family that required C grade of dexterity and strength and coordination but no minimum formal education included truck driver and appraiser, cure man, cutter, letterer, sign painter,

drill operator. These patterns of occupations might be useful in emergencies where it is necessary to recruit people in a hurry without a complete testing program. It was found, for instance, that there was greater flexibility—that is, a greater possibility of horizontal transfer—for physical labor, and less flexibility for the more skilled types of work.

SUMMARY

Job analysis involves dissecting a job both from the standpoint of the work and from the standpoint of the worker. It leads to a detailed job specification or occupational description which may be used for improving working conditions, promoting health and safety, perfecting methods of training, and supplementing employment procedure. Only the last of these is our present concern. Job specification is a necessary part of the employment program because of the ambiguity of many occupational terms and the diversity of the operations often included under a general title.

Job analysis involves much that is not psychological, but in describing the worker there is perforce a considerable use of psychological categories. Moreover, a psychological background will assist the person conducting a job analysis because of his training in observing people.

The data are usually secured by means of personal interviews with employees and executives or by observation of workers. The analyst should be familiar with the technical terminology but need not be experienced in the occupation in question. He likewise needs various personal qualities such as patience and tact. Preliminary training for interviewers is desirable and this may well take the form of trial interviews followed by criticism of the results. Similar training is desirable where the method used is observation of workers.

It is wise to interview both workers and their superiors. The former may attach little significance to acts that have become automatized, and the latter may overlook minor aspects of the job that would occur to them if they were actually performing it. In selecting persons to interview it is desirable to secure a sample that will be typical rather than one that represents only

one aspect of the job. Similar care is necessary in selecting a sample of workers for observation.

The analyst may well be provided with a work sheet calling for various items of information, such as duties, responsibilities, equipment, tools, and working conditions. Minimum requirements for the worker are to be ascertained and often rated according to a code or a brief rating scale. It is desirable that for each item, such as physical qualities or experience or judgment, the analyst give a reason for the particular entry he makes.

With these data from many interviews or observations, the occupational description or job specification can be written. It is common practice to put this in the form of simple declarative sentences with a brief paragraph covering each item, such as duties, hours, minimum qualifications, additional qualifications, working conditions, and lines of promotion. For convenient reference most of these facts may be reduced to a code notation and indicated in boxes at the top of the blank.

Many of the factors involved in job analysis are, of course, non-psychological in character. However, the analysis may be of assistance to the psychologist in initiating a project for developing tests for a given occupation. While he may have to go further in determining the mental aspects of the job for which it is advisable to develop tests, the job analysis if already conducted may well be a starting point.

On the other hand, many of the psychological methods already discussed may make some contribution to job analysis. It is possible to evaluate statistically miscellaneous items of personal history that are often included in the occupational description and are based on the judgment of those interviewed without necessarily scientific justification. The analysis frequently includes various mental traits such as are usually embodied in rating scales. If the rating scale technique is used to determine the amount of the trait necessary for the job, the applicants may be rated similarly to see whether they attain this critical amount.

Whenever the job specification calls for previous experience in a trade the desirability of a trade test is obvious. Instead of taking the applicant's word on his status, he may well be tested to determine it. The job specifications may embody a critical trade test score.

In view of what we know regarding the correlation of intelligence with vocational aptitude and the nature of the occupational hierarchy, it would seem logical for the specifications for certain jobs to contain critical scores in intelligence.

Finally, tests for special capacity may frequently be developed as a part of the job analysis procedure and critical scores embodied in the final specifications. Theoretically the job specification ought to contain everything that will promote the selection of workers who will be efficient and happy. In addition to the usual information regarding duties, hours, salary, and personal qualifications revealed by the applicant's own statements, it will in many cases promote this more effective selection if critical scores, or something analogous, are included in rating scales, trade tests, items of personal history, and tests of intelligence and special mental capacity.

REFERENCES

1. Benge, E. J. Job Evaluation in a Paper Plant. *Personnel Journal*, 1940, 19, 42-48.
2. Charters, W. W., and Whitley, I. B. *Analysis of Secretarial Duties and Traits*. Baltimore, Warwick and York, 1924, 186 pp.
3. Dye, D. E., and Unger, E. W. A Technique for Job Analysis. *Personnel Journal*, 1929, 7, 367-375.
4. Scott, W. D., and Clothier, R. C. *Personnel Management*. Chicago, Shaw, 1923, 643 pp.
5. Snow, A. J. *Psychology in Business Relations*. Chicago, Shaw, 1925, 562 pp.
6. Stead, W. H., and Shartle, C. L. *Occupational Counseling Techniques*. New York, American Book, 1940, 273 pp.
7. U.S. Employment Service. *Dictionary of Occupational Titles*. Washington: Government Printing Office, 1939; Vol. I, *Definitions of Titles*, 1287 pp.
8. Williams, R. D. How to Prepare a Job Manual. *Personnel Journal*, 1939, 18, 219-226.

Chapter XVI

THE OUTLOOK FOR PSYCHOLOGY IN SELECTING PERSONNEL

SUMMARY OF PSYCHOLOGICAL TECHNIQUE APPLIED TO EMPLOYMENT METHODS

We have now completed our survey of present-day psychological technique in so far as it bears on problems of employment. After clearing the ground of certain pseudo-psychology which is often presented to the business man as a remedy for his employment difficulties, we discussed the technique that is most widely used in this field, namely, mental tests. The distinction was made between tests of innate capacity and tests of acquired proficiency, and the former were subdivided further into tests of special capacities such as attention or memory and tests of general capacity or intelligence. Illustrations were given of a considerable variety of such tests with which an employment psychologist would ordinarily be familiar before undertaking a research project. The technique of devising and administering tests was described in detail. Attention was called to the fundamental importance of validating the tests or other measurements by comparing them with the criterion—some expression of the workers' ability in the job. It is always necessary to determine whether those who are efficient in the test are efficient in the job, and vice versa, before the tests can be validly used for occupational prognosis.

We then discussed in more detail the criterion of occupational efficiency in the form of either ratings by the employee's superiors or production figures. We also noted two possible types of subjects on whom to standardize the tests—employees and applicants. While the latter are perhaps better from the theoretical

standpoint because the tests are to be used ultimately upon applicants, nevertheless as a practical matter employees have been used more frequently as subjects in research for the simple reason that the criterion is available more quickly.

The specific procedure of validating the mental tests by comparing test scores with the criterion was next discussed. Dealing first with tests of special capacity, we saw that approach to the problem may be made through two avenues—reproducing the total mental situation involved in the job or analyzing the job into its mental components and measuring these separately. In the former case the score in the single complicated test is correlated with the criterion to determine its value. In the latter, each test is correlated separately with the criterion in order to retain the most valuable tests and discard the others. The tests in this final group are then weighted in order to allow for any overlapping of the different ones and to combine them in such a way as to secure the best possible prediction of vocational aptitude. In either instance, when the final correlation of the single test or the group of tests with the criterion is known, a critical score can be set as a basis for hiring or rejecting applicants. This critical score may best be determined by computing the probability that an applicant with a certain test score will reach a certain level of occupational success. The employment department then knows how big a chance it is taking with an applicant, and it can decide, after considering all other related factors, whether it wishes to take this chance.

We next considered general capacity or intelligence as related to vocational aptitude. Occupations appear to follow an intelligence hierarchy, inasmuch as the average intelligence of occupational groups increases consistently from unskilled labor to the professions. This suggests that a person tends to attain as high an occupation in the scale as his intelligence warrants, and gives some notion of the intellectual requirements of various occupations. In some types of work intelligence scores correlate significantly with the criterion so that the procedure used with special capacity tests is applicable. Furthermore, in some instances the work requires not necessarily a maximum intelligence but rather an optimum intelligence, for persons who are too good for their job are apt to be dissatisfied and quit.

Interest as well as ability is important in vocational prediction. Consequently, methods of measuring interest were discussed, as well as cases in which interest data were evaluated with reference to vocational success.

We then turned to the technique for dealing with certain traits, such as industry, cooperativeness, tact, and enthusiasm, which cannot at present be measured by tests but are nevertheless of vocational significance. For such traits the judgments of acquaintances or colleagues can be systematized by means of rating scales. In the man-to-man scale the person being rated is compared with others on a previously constructed master scale; in the method of defined groups he is located with reference to the distribution of similar workers into a series of groups of equal size that possess the trait in an increasing degree; in the graphic method his standing is indicated by a check mark somewhere along a line on which the rater is guided by descriptive adjectives; in the check-list method a large number of statements are checked according to whether they characterize him. In the first three methods one trait is evaluated at a time in order to abstract from errors due to general impression.

This was followed by a discussion of miscellaneous factors which may be used as a supplement to mental tests or in lieu of them if tests are not feasible. Educational status and items of personal history such as often appear on the application blank may be statistically evaluated by determining which ones are differential of occupational ability. Application letters were shown to be very unreliable; the best procedure for dealing with them is to pool the independent judgments of several persons who evaluate them. The recommendation procedure may be improved by the use of an inquiry blank calling for brief answers or check marks. The employment interview may well be supplemented by an interviewer's rating scale.

We then turned to the trade test which, instead of prophesying future occupational status, is designed to measure a person's trade skill or information at the present time. The technique consists of testing novices, apprentices, journeymen, and experts and finding which particular items or questions are differential of these groups. It is then possible to establish a critical score to determine in which of these trade classes an applicant belongs.

Finally, we discussed job analysis in so far as it bears on employment psychology. Many alleged requirements on the part of the worker can be statistically evaluated before their inclusion in the final occupational description. The rating scale technique may prove valuable in dealing with certain personality factors in the analysis. When trade experience is a necessary requirement the technique of trade tests would seem in point. In many types of work the job specification might well include critical scores in tests of intelligence or special capacity.

The foregoing are some of the psychological principles that are applicable to problems of employment. They are the result of gradual development and of the cooperation of many psychologists all along the line—from the first ones who constructed mental tests through those who perfected the statistical methods to those who have actually validated tests and other techniques in various practical fields. It now remains to look toward the future.

ATTITUDE OF WORKERS AND MANAGEMENT TOWARD PERSONNEL PSYCHOLOGY

As individual scientists and larger organizations contribute toward the advancement of personnel psychology, much depends on the attitude of those involved.

Workers. The attitude of the workers toward employment psychology has not manifested itself unmistakably as pro or con. Of course there is a natural suspicion of any innovation that apparently aims at efficiency. There have been instances when methods of scientific management have been misused, not through any fault of the principle but because of abuse of the practice. Employees have observed improvements brought about by such methods without any measurable benefit to themselves and they have naturally been disgruntled. This attitude has not extended to mental tests to any great degree. Some applicants who take such tests seem quite interested, others take them as a matter of course, and a relatively small number are annoyed, feeling that this is an undesirable method of getting a job.

An impartial consideration of the foregoing chapters will indicate that this hostile attitude is ungrounded. Employment

psychology aims to benefit the employee as well as the employer. The man who is placed in a job for which he has the aptitude will enjoy his work and will in general be happier. Applied psychology is distinctly impersonal. It aims to discover the facts and derive methods regardless of who uses them. The psychologist could just as well be retained as a consultant by the workers as by the management. Theoretically a factory operated by a council of employees should be just as interested in psychology as one run on the usual basis. It is desirable to educate the workers to realize the impersonal character of employment psychology. They should be made to see the advisability of not giving a man a job and a wage arbitrarily, but of discovering and developing his particular ability to the best advantage. There is no waste so far-reaching as misdirected human activity, and waste in industry hits all of us, including the worker himself.

The attitude toward actually taking psychological tests is reasonably satisfactory, and is becoming more so. This is largely due to their increasing use in the educational system. Most applicants for a job these days have already encountered psychological tests in school. People are coming to accept them, much as they do the medical examination, as a routine part of the procedure of getting a job.

Management. The attitude of the management toward employment psychology is likewise important. While some executives still feel self-sufficient in dealing with the human element, the majority are coming to realize their own limitations or are at least willing to submit their own opinions to scientific evaluation. They must, moreover, appreciate the scientific attitude and the necessity for investigating minutiae, for repeating observations again and again, and for amassing statistical data. They must consider the general results rather than the individual case which may be an exception to the rule. When dealing with vocational prediction it is a question of probabilities, and even though the methods are rather successful there are bound to be some erroneous predictions. Executives must learn to consider the proportion of successful placements rather than the results with a single employee. Finally, they must be patient with the slow, painstaking character of scientific research.

NECESSITY FOR FURTHER RESEARCH

Granted that workers and management are willing to cooperate in developing psychological methods for employment, it is scarcely necessary to stress the importance of further research in this field. Obviously more facts are needed, and no one can determine whether a given procedure is the proper one until it is tried out. Industries realize the importance of research in other technical lines; they hesitate to base decisions upon opinion when facts can be obtained. Many concerns, of course, maintain their own physical or chemical laboratories. Research in psychology is often just as important as in these other sciences. While a company will measure the specific gravity of certain compounds used in its products, it is less inclined to measure the mental capacities of the workers who are handling these products. It is just as impossible to solve these problems by intuition as it is to determine the weight of a liquid by looking at it.

Problems of Individual Concerns. Much of the research that is necessary grows out of the individual problems of a particular plant. Each concern frequently has its own special situations which need specific study. A technique developed in one field should not be taken bodily into another without evaluating it in the latter situation. Clerical tests developed in one plant may be unsatisfactory for use in another because computing machines are used in the first but not in the second. A rating scale developed in one organization would not necessarily work well in another, for the first concern might be rating one kind of executive and the second one a distinctly different type. Each individual concern must then validate the psychological methods in its own situation before putting them into practice. Even when tests have been fairly well standardized and put out in commercial form, it is well to make a preliminary study of them in the new situation before attaching too much value to them. Personnel men who think that it is only necessary to purchase some standard tests and begin to use them immediately for hiring are usually wrong. Any individual concern thus presents a variety of problems for psychological research.

Special Occupations. In addition to research of the above type in validating a previously developed method in a new situation

or in devising new methods for local conditions, there are other problems of a more general nature with which employment psychologists must concern themselves—problems to which many individual workers must doubtless contribute before they are finally solved. For instance, if one goes through the range of occupations he will find some in which satisfactory experimental results have been obtained and others in which apparently little has been accomplished. Considerable success has been achieved, for example, in the selection of clerical workers. This type of work apparently necessitates certain rather specific capacities that are objectively measurable by existing tests. Furthermore, in the case of clerical or industrial workers it has frequently been possible to find a considerable number of persons doing the same sort of work on whom the test can be standardized.

The situation is quite different when such complex things as executive ability are concerned. The executive has to reason, make decisions, deal with men, cooperate, get things done, delegate authority, and the like. These traits or capacities are not so readily measurable as are those required by clerical or industrial workers. Up to the present they have been approached largely through the technique of rating scales, and progress has not been rapid. More objective methods will be necessary before the problem of selecting executives is satisfactorily solved. Moreover, if the measurements themselves are perfected, it will often be difficult to validate them because it is unusual to find a considerable number of executives who do approximately the same thing. In a factory a hundred men building the same kind of automobile tire might readily be found, but a hundred executives selected from the same concern would probably be doing approximately one hundred different things, so that it would be more difficult to obtain the criterion by which to evaluate the measurements. However, as time progresses it will doubtless be possible to select certain aspects of executive ability that are rather common to a great many positions and devise methods for measuring these particular aspects.

Other occupations are in somewhat this same difficulty. Salesmanship, for instance, has been studied to quite an extent, but the problem of selecting salesmen has by no means reached its final solution. The characteristics which constitute a successful

salesman are apparently exceedingly complex, many of them involving personality traits rather than mental capacities. While various ingenious tests have been found to indicate selling success to some extent and while various items of personal history have been somewhat differential, a great deal of research still remains to be done in this field. Apparently what is needed are objective personality tests instead of the existing tests in which the subject indicates his likes, dislikes, worries, and interests, and may misrepresent them in his effort to secure the job. In the various professions there has been little research indeed. Methods for selecting professional men, except possibly engineers, are of little interest to the business man. However, the development of vocational standards for all lines of work is a step in the whole program of adjusting people more satisfactorily to the type of work for which they are best fitted.

Special Techniques. Another field for general research contribution is the development of further mental measurement techniques. Fairly satisfactory tests for some of the simpler capacities and abilities are already available. While these are undoubtedly of great importance in many occupational lines, nevertheless any psychologist realizes that other things also are necessary. It is not always a question of what a workman can do, but of what he will do. His attitude toward his work and the way he approaches it are important considerations in his occupational prognosis.

We noted some efforts to measure interest objectively. These methods when perfected will provide fairly well-standardized techniques for determining a person's vocational and avocational interests with a view to placing him in a position where these interests will facilitate rather than hinder his progress.

Then there is the whole field of temperament and personality measurement. Better methods of evaluating such things as honesty, flexibility, stick-to-it-iveness, adaptability, tact, enthusiasm, and the like are needed. At present there are systematized efforts to rate such qualities, but this usually necessitates the rater's being acquainted with the person who is to be rated. As mentioned above in connection with selecting salesmen, the need is for objective methods of measuring these qualities in the same fashion that intelligence or memory or reaction time can be measured.

Even in the field of intelligence measurements it will be recalled that three types of intelligence have been suggested—abstract, mechanical, and social. Most of the work hitherto has dealt with the first two. A field for much-needed research is the development of measurements of social intelligence in order to determine an individual's general ability in dealing with a social situation as compared with his ability in dealing with more abstract things. This technique will be especially valuable in employment problems that concern occupations in which the individual makes definite social contacts and his success in the occupation depends somewhat upon his adaptability in making such contacts.

ESSENTIALS FOR FUTURE RESEARCH

Competent Psychologist to Conduct the Research. The foregoing are some of the problems with which research workers in employment psychology must in the future concern themselves. We shall now consider some of the conditions necessary for successful research work in this field. In the first place, a competent psychologist should be obtained to conduct a given piece of research. Earlier chapters have indicated that this type of work involves rather special techniques and requires a person with some experience in mental measurements and some appreciation of individual differences. After measurements have been put into final form so that they are relatively fool-proof, it is time to turn them over to untrained individuals for routine administration. Even then there is something to be said for the value of a modicum of psychological training for those who administer tests and interpret the results. But in the process of developing methods before their final application, laboratory training is invaluable. In such a research program contingencies are apt to arise which would lead the untrained experimenter into various errors. He might fail to establish rapport, fail to control the attention of the subjects, be uncertain what to do in case of a bad start, and overlook various incidental reactions of the subjects which might be of considerable significance. No concern would put into its industrial laboratory a person who had had no chemical laboratory experience but had merely taken theoretical courses and read about chemistry. He would be likely

to drop the test tubes, mix the stoppers of the reagent bottles, and punch a hole in the filter paper. Similarly a psychologist without laboratory experience would be inclined to vary the test instructions, to overlook various conditions of illumination and the like, to fail to eliminate unnecessary distractions during the test, and to be careless with the temporal aspects of the procedure. Aside from its aid in the mere conduct of the test the laboratory background gives one a scientific attitude in interpreting the results. The uninitiated is apt to stress some aspects that appeal to him. His grading of a test blank may frequently be colored by his general impression of what the subject ought to do.

The importance of obtaining a competent psychologist is stressed because there have been instances in which business men employed persons who purported to be psychologists but who were not adequately trained. These individuals were naturally unsuccessful in their practical work and to some extent this brought discredit upon the science in general. While many of these people were not actually fraudulent, they were nevertheless incompetent and should not have been engaged in this type of work. As suggested in Chapter III, one means of ascertaining whether an individual is competent for such work is through the Psychological Corporation, which endeavors to put persons needing psychological service in contact with someone who can adequately perform that service. The directory of the American Association for Applied Psychology gives information as to the fields in which the members are experienced and lists only people who are known to be competent.

Adequate Criteria. A second essential for future employment research is adequate criteria. In Chapter VI the fact was stressed that psychological measurements can be no more valuable than the criteria by which they are evaluated. Obtaining these criteria depends on the cooperation of all those concerned in furnishing such data. If foremen or managers or others are called upon to rate the men under them in some way, it is essential that they take this work seriously and make the ratings with the greatest possible care. With reference to production criteria, of course the research depends upon full access to all production records that are available. If these records are to be valuable, obviously they must have been accurately kept. Some indication of the detailed

procedure for approaching the criterion problem was given in Chapter VI. In retail selling, for example, a dozen different items were combined to yield a criterion. In certain other operations it proved desirable to conduct two complete pieces of research, one using a quality criterion and the other a quantity criterion. The above considerations indicate the importance that is being attached to this problem of criteria.

Subjects on Whom to Standardize Methods. A third essential for such research is the subjects on whom the experiments are to be conducted. Access must be had to employees (or possibly applicants) on whom to standardize the various measurements. The psychologist must go into the plant with his tests and measurements. He could not, for instance, standardize a vocational test for lathe operators on students in an Arts College; he must evaluate it with men who are actually doing the practical work. This may cause some inconvenience at the plant where the research is being done, but it is nevertheless necessary. Furthermore, the subjects who are used must cooperate and do their best in taking the tests. The only way to keep incentive constant, as has already been suggested, is to keep it at a maximum. Results will naturally be meaningless if one subject does his best and another does not try. Consequently, such a program cannot be carried through successfully where the morale is low and the persons taking the tests are unwilling to cooperate. In addition to obtaining employees who are willing to do their utmost, it is further necessary to have enough of them to make the results reliable. The psychologist cannot be expected to solve the problems of a given vocation by having six men sent to him for testing. According to the general principle of averages, the more that are included the more likelihood that the results will represent typical tendencies.

Facilities for Conducting Research. A fourth essential for personnel research is adequate facilities for conducting the work. In giving a test, for instance, it is essential that all the subjects have approximately standard conditions. It would be impractical to test some of them in the shop and some in the laboratory because of the different amount of distraction. A separate laboratory is desirable, for lighting, ventilation, and other external conditions can be kept in an optimum condition. Adequate time

should be allowed, moreover, for each subject who is tested. If it becomes necessary to rush, the examiner is likely to make various errors himself and his attitude of excitement is quite apt to be communicated to the subjects. Consequently, no psychologist would be enthusiastic about testing a group of men at lunch hour or after the day's work. It is usually necessary to test the men during working hours on the company's time in order to insure standard conditions and adequate time as well as proper morale. Obviously it is desirable to convince the company of the importance of the above factors before undertaking the personnel research program.

Opportunity to Evaluate Results Adequately. A fifth requisite is an opportunity to study the results adequately without pressure. An executive is likely to consider a psychologist as he does his salesman and look for immediate returns. It is unwise to crowd a research worker. Discoveries cannot be made to order. Hence the management must be patient with the research department. Sometimes the research workers naturally go into blind alleys and must start over again. But if one considers the number of reagents that were tried before the discovery of the one which when mixed in gasoline eliminates the knock, he will be inclined to pardon an employment psychologist for making a few false starts that do not lead directly to the mark. Scientific facts do not spring up overnight. A condition that greatly irritates a research worker is pressure to uncover fundamental truths on schedule. In this connection he should have ample opportunity to follow up his results. He may have devised a set of measurements which apparently indicate aptitude for a particular line of work, but he may not be fully satisfied with the results until he has checked them on a new group of people who are selected on the basis of these measurements and who subsequently demonstrate their fitness or unfitness. This subsequent validation of the measurements should by all means be not only permitted but encouraged.

General Cooperation. Finally, the research worker requires the general cooperation of all those with whom he comes in contact. The scientist will not do his best if someone is continually opposing him. His own morale should be considered as well as that of the workers. He often needs advice on many points, he requires

records and supplies, clerical assistance is sometimes necessary, and various accommodations may have to be made for him in shifting schedules or providing as subjects a particular group of workers who are of special interest. Everyone with whom he is working must be definitely "with him" in the project. It is preferable for him to be considered, temporarily at least, as an integral part of the staff or at least to have his status in the organization definitely recognized.

THE SOCIAL IMPLICATIONS OF EMPLOYMENT PSYCHOLOGY

Before concluding the discussion of the outlook for employment psychology, we should consider again its broad social implications brought out at the end of the first chapter. The methods described in this book can be of as much benefit to the employee as to the employer. It is really a kindness to an applicant not to hire him for a job in which he has little chance of success, because there is thus a greater probability that he will locate something in which he has a future. Misdirected human activity is one of the greatest wastes in our civilization and it indirectly affects all of us. Furthermore, while the techniques discussed above are for the most part objective, impersonal, and statistical, this does not mean that the employment process should be stereotyped and mechanical. As suggested in Chapter I, the applicant must after all be regarded as an individual who has certain capacities and likewise certain interests and who is looking for opportunities. His interests must be treated with respect and tact, especially if they are apparently at variance with his capacities. The employment technique should be tempered with a certain amount of common sense and appreciation of the unique problems of the individual. He should be aided so far as possible in finding himself and in improving his opportunity. But even after these things are taken into consideration, the major part of the problem still consists in measuring the man's innate potentialities and comparing them with objective standards that have been developed for the particular job in question. This is the largest contribution which psychology can make to increasing human efficiency by scientific selection of personnel.

Efficiency, however, should not be achieved at the expense of happiness nor should happiness be obtained at the expense of

efficiency. The happiness to be considered, however, is ultimate rather than immediate—the real satisfaction that comes from the expression of normal cravings for achievement, freedom from fear or jealousy, reasonable leisure, and a sense of accomplishing something worth while. From this standpoint the capacities and the interests of the man should be considered and the attempt made to adapt him to his work and adapt the work to him so that this unit will have maximum effectiveness. Employers often shy at the idea of happiness as one of the goals for scientific effort. Some of them doubtless have had unfortunate experiences with professional uplifters. But the psychologist is not thinking in these terms, and he is not shortsighted in his belief that a happier society is a more effective society. It is difficult to say how much of our industrial unrest and unhappiness is due to the maladjustment of the worker to his work. What is often stated as the cause of the unrest is often not the real cause. In many cases persons have been known to protest about their wages when what was really bothering them was the climate. They apparently find disagreeable aspects in their working conditions when the real trouble is that they are individually not adapted to their work.

In this scheme of things applied psychology will in the future play an increasingly large role. The science has, to be sure, been "oversold" in a few instances. It is a rather common tendency among business people and others to claim too much for something they have to sell, and psychology is no exception, for its surplus enthusiasm at one time led it a little too far in this respect. But the lean years of the business cycle purified its soul. It has gone back again to fundamentals and is proceeding with painstaking and thorough scientific procedure.

Personnel research is a comparatively new study, mental measurement is not familiar to and appreciated by the layman, and it will take considerable time before people come to appreciate these things fully. It took a long time, for instance, to remodel our social attitude toward crime. The same thing will doubtless be true of the social attitude toward applied psychology in general and employment psychology in particular.

The broad movement to study man has just begun. Psychology is now playing an increasing role in the schools, in the clinic, in

the advertising agency, in the factory, and in the employment office. These problems of life adjustment are coming more and more to the front. The last century was characterized by tremendous advances in the natural sciences and in the technologies. The present one bids fair to be an era of human engineering. The psychologist's ideal is to have everyone provided with the opportunity to do that particular part of the world's work for which he is best adapted and in which he is most interested. When this ideal is achieved, the world will be a better place for all of us.

Appendix 1

ILLUSTRATING THE TECHNIQUE OF CORRELATION

The idea of correlation is fundamental in personnel psychology. We are often concerned with the extent to which two variables or sets of traits or measurements are related. We may wish to determine whether or not estimates of a trait made by acquaintances are at all related to estimates made by unacquainted persons who judge only from physiognomy. We may desire to ascertain to what extent efficiency in a particular mental test is related to efficiency in a job. The ultimate aim is usually to predict one variable in terms of another; hence the need for expressing quantitatively the relation between the two variables. The correlation coefficient is the standard technique for expressing this relation. The present section aims merely to give a simple notion of how correlations are obtained and the meaning of correlations of different magnitudes. The examples cited are made absurdly brief in the interest of avoiding tedious arithmetical computations. With longer examples, the arithmetical work would naturally be more arduous, but various short-cut procedures are available.

One of the simplest correlation procedures is that involving rank differences. Given two series of measures, it is possible to rank them both and obtain the differences in the rank. Consider Example I, which gives data for five men. Let us suppose that some quantitative statement of their ability in the job, such as units of production, gives the scores indicated in the first column of figures, and these men make the scores in a mental test indicated in the second column. The problem is the extent to which those who make high test scores make high job scores, and vice versa.

It is to be noted that in job scores Briggs is the best of the group and is ranked 1 (cf. the column headed "job rank"); Andrews is second best and gets a rank of 2, and Adams falls in third place. Similarly, Adams makes the highest test score and is given a rank of 1 (cf. column headed "test rank"); Andrews is next best and gets

EXAMPLE I

Name	Job Score	Test Score	Job Rank	Test Rank	Rank Difference	Rank Difference Squared
Adams.....	75	45	3	1	2	4
Andrews.....	80	43	2	2	0	0
Briggs.....	82	38	1	3	2	4
Brown.....	63	36	5	4	1	1
Doc.....	68	34	4	5	1	1
					—	10

Sum of rank differences squared = 10

$$\rho = 1 - \frac{6\sum D^2}{N(N^2 - 1)} \qquad \sum D^2 = 10 \qquad N = 5$$

$$\rho = 1 - \frac{6 \times 10}{5(25 - 1)} = 1 - \frac{60}{5 \times 24} = 1 - .50 = .50$$

a rank of 2, and Briggs comes third. We may neglect the first two columns and, considering only the columns of ranks, determine the difference in rank in each instance. Adams is ranked 3 on the job and 1 on the test; the difference between these figures is 2 (cf. the column headed "rank difference"). Andrews is ranked 2 in both cases so the difference is 0. These differences give some notion as to the extent to which the two series of ranks correspond. If the difference is as great as 4 or 5 it indicates that a person is ranked in one trait very differently from the way in which he is ranked in the other. If the difference is small it indicates that there is a fair correspondence.

In working out the correlation coefficients on the basis of these data it is necessary to square the rank differences, as is done in the last column. The sum of these squares is then obtained. The formula for computing the coefficients is indicated in the first example. The coefficient is designated by the Greek letter rho. The formula must be taken on faith in the present connection, but its derivation can be obtained in advanced works on statistics. In the formula the term $\sum D^2$ means the sum of the squares of the differences, while N means the number of cases involved—in this instance 5 men. To solve the formula it is necessary to take 6 times the sum of the differences

squared, divide this by N times $N^2 - 1$, and subtract the quotient from 1. In the present example this works out to an answer of 50. It is conventional procedure to carry correlation coefficients to two decimal places. This particular coefficient indicates a fair degree of correlation which, of course, is obvious from inspection of the original data, but it would not be so obvious if a large number of individuals had been involved.

For purposes of comparison several other examples are presented using similar data that give higher or lower correlations than those in Example I. Example II, for instance, is a case of what is termed perfect correlation. Briggs, who ranks highest in the job, is also

EXAMPLE II

Name	Job Score	Test Score	Job Rank	Test Rank	Rank Difference	Rank Difference Squared
Adams.....	74	38	3	3	0	0
Andrews.....	79	43	2	2	0	0
Briggs.....	82	45	1	1	0	0
Brown.....	68	36	4	4	0	0
Doe.....	62	33	5	5	0	0
						—
						0

$$\sum D^2 = 0$$

$$N = 5$$

$$r = 1 - \frac{6 \times 0}{5(25 - 1)} = 1 - \frac{0}{5 \times 24} = 1.00$$

highest in the test. Andrews, who is second in the job, is second in the test, and so on down to Doe, who is poorest in each respect. In this case there are no differences in rank and the correlation coefficient comes out 1.00, which is the maximum possible. This indicates a perfect correspondence between the two variables.

Example III presents a negative correlation. Doe, who is best in the job, is worst in the test; Brown, who is second best in the job, is second worst in the test, and so on down to Adams, who is worst in the job and best in the test. This, of course, makes the differences in rank as large as possible and the coefficient is -1.00 , which is the maximum possible negative coefficient. It indicates a perfect tendency

EXAMPLE III

Name	Job Score	Test Score	Job Rank	Test Rank	Rank Difference	Rank Difference Squared
Adams.....	64	46	5	1	4	16
Andrews.....	67	43	4	2	2	4
Briggs.....	75	38	3	3	0	0
Brown.....	78	36	2	4	2	4
Doe.....	83	34	1	5	4	16
						<hr/> 40

$$\Sigma D^2 = 40$$

$$N = 5$$

$$r = 1 - \frac{6 \times 40}{5(25 - 1)} = 1 - \frac{240}{5 \times 24} = 1 - 2 = -1.00$$

for the highest scores in one variable to go with the lowest scores in the other.

Example IV involves a correlation of .80, which is not perfect although very high. There are slight discrepancies in rank, enough to spoil the perfection; but from inspection it is obvious that there is a very close relation between the variables and this is reflected in the high coefficient.

Example V indicates a 0 correlation—that is, a situation in which there is no apparent relation between the two variables. Inspection reveals that it would be practically impossible to predict a man's job rank if his test rank were known. This is reflected in the coefficient of 0.

The method of rank differences, while a convenient and relatively easily computed correlation procedure, is not ideal because it assumes that the differences between any two adjacent ranks in one variable are all equal. Referring to Briggs' job score in Example I, we see that it is actually 2 points better than Andrews' while the latter is 5 points superior to Adams', but in ranking them it is assumed that these differences are equal. In this way a striking superiority or inferiority of some individual may be overlooked. A standard method is available which takes into account the actual magnitude of the scores. Instead of considering merely whether a person who is best in the test is

EXAMPLE IV

Name	Job Score	Test Score	Job Rank	Test Rank	Rank Difference	Rank Difference Squared
Adams.....	75	40	3	3	0	0
Andrews.....	79	48	2	1	1	1
Briggs.....	85	44	1	2	1	1
Brown.....	68	32	4	5	1	1
Doe.....	64	37	5	4	1	1
						4

$$\Sigma D^2 = 4$$

$$N = 5$$

$$\rho = 1 - \frac{6 \times 4}{5(25 - 1)} = 1 - \frac{24}{5 \times 24} = 1 - .2 = .80$$

best in the criterion, we are concerned with whether a man who deviates from the average in one respect deviates correspondingly in the other.

EXAMPLE V

Name	Job Score	Test Score	Job Rank	Test Rank	Rank Difference	Rank Difference Squared
Adams.....	86	43	1	2	1	1
Andrews.....	73	48	4	1	3	9
Briggs.....	68	37	5	4	1	1
Brown.....	77	39	3	3	0	0
Doe.....	80	31	2	5	3	9
						20

$$\Sigma D^2 = 20$$

$$N = 5$$

$$\rho = 1 - \frac{6 \times 20}{5(25 - 1)} = 1 - \frac{120}{5 \times 24} = 1 - 1 = .00$$

EXAMPLE VI

	Crite- rion	Test Score	Deviations		Deviations Squared		Product of Devi- ations	From Equation if $r = .87$			From Equation if $r = 0$		
			Crite- rion	Test	Crite- rion	Test		Crite- rion Forecast	Error of Forecast	Error Squared	Crite- rion Forecast	Error of Forecast	Error Squared
	X	Y	x	y	x^2	y^2	xy						
Adams.....	15	10	+6	+4	36	16	+24	13.4	1.6	2.56	9	6	36
Andrews.....	8	7	-1	+1	1	1	-1	10.1	2.1	4.41	9	1	1
Briggs.....	6	5	-3	-1	9	1	+3	7.9	1.9	3.61	9	3	9
Brown.....	6	2	-3	-4	9	16	+12	4.6	1.4	1.96	9	3	9
Doe.....	10	6	+1	0	1	0	0	9.0	1.0	1.00	9	1	1
Total.....	45	30			56	34	+38			13.54			56
Average.....	9	6			11.2	6.8				2.71			11.2
Sq. root of av.					3.34	2.61				1.64			3.35

$$r = \frac{\sum xy}{N\sigma_x\sigma_y} = \frac{38}{5 \times 3.34 \times 2.61} = \frac{3.34}{2.61}$$

$$r = 5 \times \frac{38}{3.34 \times 2.61} = \frac{38}{43.6} = .87$$

$$X - 9 = \frac{.87 \times 3.34}{2.61} (Y - 6)$$

$$X = 1.11 (Y - 6) + 9$$

$$= 1.11 Y + 2.34$$

$$\frac{3.35 - 1.64}{3.35} = 51\%$$

Example VI illustrates the computation of correlation by the so-called "products-moments" (that is, products of deviations) method. The first part of the computation is identical with that previously described in connection with standard deviation (p. 191). The original scores are given in the first two columns. Conventional procedure calls these original scores X and Y . The average of each column is computed. The third column gives the deviations of each criterion score from the average. The deviations are denoted by x and y . Adams' score of 15 is 6 greater than the average of 9, that is, its deviation is +6; Andrews' criterion score of 8 is one below the average, Briggs' is 3 below, etc. Deviations of test scores from the average are computed similarly. The deviations are now squared and each column is averaged. The square roots of these averages give standard deviations of 3.34 and 2.61 respectively. σ_x denotes the standard deviation of the criterion scores, and σ_y the standard deviation of the test scores.

The next step is to take the product of the deviations. For instance, Adams' deviation of +4 in the test is to be multiplied by the corresponding deviation of +6 in the criterion, giving a product of 24. Andrews' figures are +1 and -1 and the product is -1. These products are then algebraically totaled, giving 38, and we are ready to substitute in the formula. Σxy denotes the sum of the products of the deviations of the variables—in this case 38; N denotes the number of individuals, and σ_x and σ_y the standard deviations of the two variables as above described. Substituting in the formula, we obtain the coefficient of .87. This has taken into account the actual magnitude of the original measures and not merely their relative standing.

It is to be noted that if a given individual's measures are both above or below the average, the product of the deviations will be plus and the numerator of the fraction in the formula large, while if one is above the average and the other below, the product will be negative and the sum of the products will be somewhat decreased and the coefficient lowered. This type of coefficient gives probably the best indication of the relation between the two measures and is widely used in the most careful statistical work. The remainder of the table deals with the interpretation of correlation coefficients and will be discussed in a moment.

The above example is misleading as to the simplicity of the arithmetical work because correlation coefficients are never computed with only five cases. When 50 or 100 individuals are involved, the arithmetical work becomes arduous if the methods just described are used. There are, however, various short-cut procedures, but this is not the place to present them in detail. For example, the correlation

formula can be transformed so that it is expressed in terms of gross scores (denoted by capitals instead of small letters) rather than deviations as used in Example VI. Such a formula is:

$$\frac{N\Sigma XT - \Sigma X\Sigma T}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \sqrt{N\Sigma T^2 - (\Sigma T)^2}}$$

When calculating machinery is available, this gross score formula is comparatively simple because the products of the scores can be accumulated rather readily on the machine even though the products involve three or four digits. With certain types of machines it may be possible to take several of these steps simultaneously. In large-scale investigations punched-card machines can be used.

Another common practice is to group the data into class intervals rather than using the original gross scores. After they are so grouped they can be tabulated in a scatter plot (p. 231). Each row and column may be numbered from the small to the large end of the distribution and thus comparatively small numbers be used in computing standard deviations and products of deviations. With one procedure, instead of the XY products being obtained for each individual cell of the scatter plot, the entries along each diagonal can be totaled; the following formula, which has been transformed accordingly, is then used:

$$\frac{\frac{N}{2} [\Sigma X^2 + \Sigma T^2 - \Sigma (X - T)^2] - \Sigma X\Sigma T}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \sqrt{N\Sigma T^2 - (\Sigma T)^2}}$$

Short-cut procedures are discussed in current statistical works.¹

The six columns at the right in Example VI are concerned with the interpretation of correlation coefficients in terms of the error involved in predicting the criterion from the test (p. 29). The regression equation for expressing the criterion X in terms of the scores in the test Y (p. 233) is derived below the table. It is possible for purposes of discussion to work backward and predict the criterion on the basis of the test by using this equation just as if the criterion were not known. For instance, for Adams, Y is 10 and substituting this in the equation gives 13.4 as the criterion forecast or prediction. This appears in the column headed "From Equation if $r = .87$." The criterion is similarly forecast for the other workmen. The next column gives the actual error of forecast, i.e., the difference between the predicted and the

¹ H. E. Garrett, *Statistics in Psychology and Education*. New York, Longmans, Green, 1937, 493 pp.; H. A. Toops, *Computational Statistics* (mimeographed).

actual criterion. For example, the difference between 13.4 and 15 is 1.6. We then want some notion as to the average error of forecast; the standard procedure for averaging errors is to square them, average them, and take the square root. This is shown in the next column, 1.64 being the average error of the forecast. In the next three columns the same thing is shown where there is no correlation whatever between test and criterion—in other words, where we simply guess at the criterion. Obviously the only forecast we would make here would be to guess that everyone would be average, that is, have a criterion score of 9. Computing the errors of these as deviations from the actual criterion and obtaining the average gives 3.35. Thus, by using the test, we reduced the error from 3.35 to 1.64, a reduction of 51 per cent. This is the type of consideration that was used in deriving Table 2 (p. 30). It is also possible to use a general formula to compute the extent to which we can reduce the error of prediction with any given correlation. The formula is:

$$1 - \sqrt{1 - r^2}$$

In the present case, substituting .87 in the equation gives 51 per cent. The figures in Table 2 were derived in this manner.

Appendix II

ILLUSTRATING THE DERIVATION OF A REGRESSION EQUATION

In Chapter IX we saw the importance of partial correlations and the regression equation for weighting a number of tests in securing the best possible prediction of vocational aptitude. A brief indication of the technique was given in that connection. In the present section a regression in four variables is worked out in detail to illustrate the process. This is done by what is essentially the long method to show what is involved in the procedure. As in most statistical techniques, various short cuts are available to simplify the work greatly.

The equation used in the present example is the one already mentioned (p. 266) for predicting ability at finishing tires. The original correlations of the tests with the criterion and with each other are as follows, where 1 is the criterion and 2, 3, and 4 are the tests.

	2	3	4
1	.51	.49	-.41
2		.66	-.24
3			-.22

For instance, r_{12} —that is, the correlation between the criterion and Test 2—is .51; r_{13} is .49, and r_{34} is -.22. From these correlations, which are termed “zero order” coefficients, it is possible to derive any coefficient of the “first order” like $r_{12.3}$, which means the correlation between the criterion and Test 2 with Test 3 kept constant. Such coefficients are called “first order” because there is one secondary subscript, that is, one subscript after the point or one variable that is kept constant. There are always two primary subscripts before the point. From coefficients of the first order it is possible to derive those of the second order like $r_{12.34}$, which have two secondary subscripts or two variables that are kept constant. From these may be derived third order coefficients such as $r_{12.345}$, etc.

The formula used in all such computations is of the form:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2} \sqrt{1 - r_{23}^2}}$$

The subscripts of the first term in the numerator, it is to be noted, are the same as the primary subscripts of the coefficient for which we are solving (12). The second term in the numerator is the product of two factors. Each of these has a subscript (3) the same as the secondary subscript of the coefficient for which we are solving. The primary subscripts of this coefficient for which we are solving appear also in this second term, one in each factor. Putting it in another way, we obtain the subscripts of this second term by combining the secondary subscript of the coefficient for which we are solving (3) first with one of its primaries (1) and then with the other (2). The two subscripts that appear in the denominator of the formula are identical with those in the second term of the numerator.

If in this formula we substitute the zero order coefficients given in the table above we have:

$$\begin{aligned} r_{12.3} &= \frac{.51 - .49 \times .66}{\sqrt{1 - .49^2} \sqrt{1 - .66^2}} = \frac{.51 - .324}{\sqrt{.759} \sqrt{.564}} \\ &= \frac{.186}{.87 \times .75} = \frac{.186}{.653} = .28 \end{aligned}$$

This tells us that the correlation between the criterion and Test 2 would be .28 if we had persons with identical ability in Test 3. In exactly the same way the other coefficients of the first order may be derived. For instance:

$$r_{24.3} = \frac{r_{24} - r_{23} r_{34}}{\sqrt{1 - r_{23}^2} \sqrt{1 - r_{34}^2}}$$

Here again it is to be noted that the subscripts of the first term in the numerator are the primary subscripts of the coefficient for which we are solving (24); these same primaries appear in the second term—one in each factor—while the secondary subscript (3) appears in both factors and the subscripts in the denominator are the same as those in the second term in the numerator. Substituting the zero order coefficients in this formula, we have:

$$\begin{aligned} r_{24.3} &= \frac{-.24 - .66 \times (-.22)}{\sqrt{1 - .66^2} \sqrt{1 - (-.22)^2}} = \frac{-.24 - (-.145)}{\sqrt{.564} \sqrt{.952}} \\ &= \frac{-.095}{.751 \times .975} = \frac{-.095}{.732} = -.13 \end{aligned}$$

In this manner all the coefficients of the first order can be computed.

In the present problem they are not all necessary, but those that are required for subsequent use are as follows. Their method of derivation is identical with the preceding.

$r_{12.3}$.28	$r_{14.2}$	-.34
$r_{12.4}$.47	$r_{14.3}$	-.36
$r_{13.2}$.24	$r_{24.3}$	-.13
$r_{13.4}$.45	$r_{34.2}$	-.08

From these coefficients of the first order we may now compute those of the second order. The formula is similar in form to the preceding except that we are expressing a coefficient with two secondary subscripts in terms of coefficients with one secondary subscript.

$$r_{14.23} = \frac{r_{14.2} - r_{13.2} r_{34.2}}{\sqrt{1 - r_{13.2}^2} \sqrt{1 - r_{34.2}^2}}$$

The similarity of this formula to the preceding is obvious. The secondary subscript is the same throughout (2). The primary subscripts of the first term in the numerator are the same as the primary of the coefficient for which we are solving (14). The secondary subscript of the coefficient for which we are solving that does not appear as a secondary in the numerator (3) appears in both primaries in the second term of the numerator. The primaries (14) of the coefficient for which we are solving also appear as primaries in the second term in the numerator—one in each factor. The subscripts in the denominator are the same as those in the second term of the numerator. Substituting the proper values in this formula, we have:

$$\begin{aligned} r_{14.23} &= \frac{-.34 - .24 \times (-.08)}{\sqrt{1 - .24^2} \sqrt{1 - (-.08)^2}} = \frac{-.34 - (-.019)}{\sqrt{.942} \sqrt{.994}} \\ &= \frac{-.34 + .019}{.971 \times .994} = \frac{-.321}{.968} = -.33 \end{aligned}$$

This same coefficient can be computed by another formula as a check.

$$r_{14.23} = \frac{r_{14.3} - r_{12.3} r_{24.3}}{\sqrt{1 - r_{12.3}^2} \sqrt{1 - r_{24.3}^2}}$$

This conforms to the specifications mentioned in explaining the other formula for $r_{14.23}$, only it uses a different set of first order coefficients. In this case the secondary subscript that appears throughout is 3 instead of 2. Substituting, we have:

$$\begin{aligned} r_{14.23} &= \frac{-.36 - .28 \times (-.13)}{\sqrt{1 - .28^2} \sqrt{1 - (-.13)^2}} = \frac{-.36 - (-.036)}{\sqrt{.922} \sqrt{.983}} \\ &= \frac{-.36 + .036}{.960 \times .991} = \frac{-.324}{.951} = -.34 \end{aligned}$$

This checks approximately with the result of the other formula. The difference of .01 is due to the fact that the coefficients of the first order were computed merely to two decimal places. If further decimals had been retained, the check would theoretically be perfect. If the proper coefficients of the first order are available, it is possible to compute all those of the second order in two ways to detect any mistakes in the work up to that point. Making similar computation for the other coefficients that are necessary in the present problem, we have:

$r_{12.34}$.25
$r_{13.24}$.23
$r_{14.23}$	-.33

Before we can compute the regression equation we need to know the average (or mean) of each variable as well as its standard deviation (σ). These figures are as follows:

	MEAN	σ
X_1	.00	.72
X_2	28	10
X_3	19	6
X_4	210	15

The notation X_1 indicates original scores in the criterion; they were so arranged that their mean was 0. Their standard deviation was .72. X_2 denotes score in Test 2. Its average was 28 and its standard deviation 10, etc. The formula for the regression equation is:

$$x_1 = \frac{\sigma_{1.234}}{\sigma_{2.134}} r_{12.34} x_2 + \frac{\sigma_{1.234}}{\sigma_{3.124}} r_{13.24} x_3 + \frac{\sigma_{1.234}}{\sigma_{4.123}} r_{14.23} x_4$$

The r factors are the partial correlation coefficients computed previously. The x values represent deviations of a particular measure from the mean of that measure, x_2 , for instance, indicating the deviation of a measure from the mean score in Test 2. The σ values, which represent the standard deviation of a variable with the effect of the others eliminated, must be computed thus:

$$\begin{aligned}\sigma_{1.234} &= \sigma_1 \sqrt{1 - r_{12}^2} \sqrt{1 - r_{13.2}^2} \sqrt{1 - r_{14.23}^2} \\ \sigma_{2.134} &= \sigma_2 \sqrt{1 - r_{23}^2} \sqrt{1 - r_{24.3}^2} \sqrt{1 - r_{23.24}^2} \\ \sigma_{3.124} &= \sigma_3 \sqrt{1 - r_{23}^2} \sqrt{1 - r_{34.2}^2} \sqrt{1 - r_{13.24}^2} \\ \sigma_{4.123} &= \sigma_4 \sqrt{1 - r_{34}^2} \sqrt{1 - r_{24.3}^2} \sqrt{1 - r_{14.23}^2}\end{aligned}$$

The first factor in each product is the ordinary standard deviation of the variable whose number appears as the primary subscript at the left side of the equation. In $\sigma_{1.234}$ the first subscript is of the zero order

(12); the next one is of the first order and is obtained by putting the 2 into the secondary and bringing in another primary (3); for the last factor the 3 goes over into the secondary and the remaining one (4) is brought into the primary. It is to be noted that 1 appears as a primary throughout. The other formulae embody the following principles: The subscript that appears as the primary at the left of the equation remains as a primary throughout. The last factor always has 1 as one of its primaries. Hence the last factor can always be determined by using as primaries 1 and the primary that appears at the left of the equation and using all the other variables as secondaries. The next to the last factor is obtained by shifting one of the secondaries into the primary, displacing the primary that is not to be a primary throughout. For instance, take $\sigma_{4.123}$. The last factor must have 1 as a primary subscript and also 4 which appears at the left of the equation as a primary. The next to the last factor drops 2 from the secondary and puts it in the primary to replace 1. It cannot replace 4 because that must remain as a primary throughout. The first factor now has this secondary 3 dropped and moved into the primary, replacing 2. Substituting the appropriate values in these equations gives:

$$\sigma_{1.234} = .72\sqrt{1 - .51^2}\sqrt{1 - .24^2}\sqrt{1 - (-.33)^2} = .72 \times .862 \times .971 \times .944 = .573$$

$$\sigma_{2.134} = 10\sqrt{1 - .66^2}\sqrt{1 - (-.13)^2}\sqrt{1 - .25^2} = 10 \times .751 \times .991 \times .968 = 7.15$$

$$\sigma_{3.124} = 6\sqrt{1 - .66^2}\sqrt{1 - (-.08)^2}\sqrt{1 - .23^2} = 6 \times .751 \times .997 \times .973 = 4.38$$

$$\sigma_{4.123} = 15\sqrt{1 - (-.22)^2}\sqrt{1 - (-.13)^2}\sqrt{1 - (-.33)^2} = 15 \times .975 \times .991 \times .944 = 13.68$$

We are now ready to substitute in the regression equation.

$$x_1 = \frac{.573}{7.15} .25 x_2 + \frac{.573}{4.38} .23 x_3 - \frac{.573}{13.68} .33 x_4 \\ = .02 x_2 + .03 x_3 - .014 x_4$$

There is one further step to take before the equation is in its most useful form. As above given, it involves the deviations of the scores from their mean rather than the actual original scores. If it were to be used in this form, it would be necessary to convert each measure into a deviation and substitute in the equation; then, after x_1 had been obtained, to convert it back into terms of actual score. It is better to make a single transformation for the whole equation so that original scores can be substituted in it directly. This can be done by virtue of the fact that a deviation is simply the original score minus the mean, so that $x_1 = X_1 - M_1$, $x_2 = X_2 - M_2$, etc., where x_1 is the deviation of the criterion, X_1 the original score, and M_1 the mean of the criterion scores, and the same meaning is attached to x_2 , the deviation of a score

in Test 2, etc. The mean scores have been given above so that we may make the following substitutions:

$$\begin{aligned}x_1 &= X_1 - 0 \\x_2 &= X_2 - 28 \\x_3 &= X_3 - 19 \\x_4 &= X_4 - 210\end{aligned}$$

Making this transformation, we have:

$$\begin{aligned}X_1 - 0 &= .02 (X_2 - 28) + .03 (X_3 - 19) - .014 (X_4 - 210) \\X_1 &= .02 X_2 + .03 X_3 - .014 X_4 + 1.82.\end{aligned}$$

This is the final form of the regression equation, and if a given applicant has taken the three tests his scores can be substituted in this equation to obtain his most probable score in the criterion. (Cf. p. 267.)

When more than four variables are involved, the labor of computing the coefficients increases, but the procedure above outlined does not have to be followed, for various short cuts are available. Even so, the technique of partial correlation is tedious, but worth while.

Instead of determining in advance just how many tests are to be used and working out all the necessary partial correlations, it is possible to add tests to the battery one at a time in the order of their importance and at each step determine the multiple correlation, that is, the correlation with the criterion of the total weighted score of the tests thus far included. When the point is reached at which an additional test makes no significant increase in the multiple correlation it is useless to go further. As an additional refinement of this approach, a correction for the chance error added by the test may be applied at each step. As more tests are added to the equation the chance error usually increases and the increment to the multiple correlation becomes less and less. When finally the point is reached at which the addition of another test adds more chance error than actual validity to the battery, it is time to stop. A widely used procedure of this sort is the Wherry-Doolittle method.¹ This method applied to the data in the above example yields a multiple correlation of .59, which is slightly smaller than the multiple of .61 obtained by the original method. This difference probably is due to the addition of some chance error by the various tests.

¹ W. H. Stead and C. L. Shartle, *Occupational Counseling Techniques*. New York, American Book, 1940, Appendix 5.

INDEX

-
- Ability and interest, 322
 Academic record, vocational prediction from, 406 ff.
 Accuracy vs. speed scores, 158
 Acquaintance, effect on ratings, 391
 Acuity, tests of, 80
 Administration of tests, 122 ff.
 Age, job satisfaction, 425; test score differences, 207; turnover, 424; vocational success, 422
 Aged, test performance, 210
 Aggressiveness tests, 115
 Agricultural engineering, interest test, 332
 Alpha, Army, 58
 Alternative answer tests, 129
 Alternative answers, location of, 141; selection of, 140
 Alternative test material, 143
 American Association for Applied Psychology, 18
 American Psychological Association, 17, 65
 Analysis of job for selecting tests, 221 ff.
 Apparatus, improvised, 225
 Applicants as subjects, 201
 Application letter, 436 ff.
 Applied psychology, late development, 2
 Aptitude vs. achievement tests, 73
 Army Alpha test, 58
 Association tests, 85
 Astrology, 18
 Attempts vs. total test score, 158
 Attention tests, 82 ff.
 Attitude of person being tested, 136
 Automobile drivers, tests for, 216, 272
 Binet, 55, 106
 Blonde and brunette traits, 41
 Building trades, tests, 269
 Capacity vs. proficiency tests, 73
 Cashiers, intelligence of, 311
 Character analysis, unscientific methods of, 16 ff.
 Character Education Inquiry, 113
 Check-list rating scale, 375
 Clerical tests, 270 ff.; factor analysis, 271
 Clerical work and intelligence, 300
 Clothing operators, intelligence of, 302
 Color blindness, 80
 "Completion" test, 130
 Components, mental, of job, 217, 254 ff.
 Conversation as measure of interest, 337
 Cooperative research, 63
 Coordination, tests, 77
 Correlation, criterion and test, 228 ff.; interpretation of, 29; partial, 60, 162, 259, 560; products-moments, 230, 555; rank difference, 229, 551; scatter plot, 230
 Cotton mill, tests, 269
 Criterion, availability, 172, 203; combining different types, 188; foreman's estimates, 179; importance, 171; length of service, 187; linear scale estimates, 182; miscellaneous types, 186; predicted from regression equation, 239; production, 173; quality vs. speed, 268; reliability, 178, 185; speed vs. quality, 268; supervisor's estimates, 179; for trade tests, 484; weights of variables, 196
 Critical score, 236; graphic method, 249; intelligence, 306; upper, 313
 Cycle omnibus test, 143
 Deception tests, 117
 Dexterity tests, 77

- Dial machine operators tests, 225
 Difference, significance of, 328;
 standard deviation of, 329
 Driving clinic, 272
- Early success in job predictive of
 later, 415 ff.
- Elderly people, test performance,
 210
- Electrical manufacturing, tests, 269
- Emotion tests, 116
- Employees as subjects, 202
- Employment Service, United States,
 66
- Executives, intelligence, 302; tests,
 277
- Experience, as influencing criterion,
 176, 203; predictive of vocational
 success, 428
- Factor analysis, of aptitude tests,
 278; of clerical tests, 271; of in-
 telligence tests, 107; of interests,
 330; technique, 107
- Factory operatives, tests for, 268 ff.
- Faculty vs. student intelligence, 291
- "Families" of jobs, 531
- Feeble-minded, vocations for, 297
- Firemen tests, 273
- Follow-up procedure, aptitude tests,
 280
- Fool-proof tests, 226
- Forecasting efficiency of correlation
 coefficient, 30
- Foreman's estimates as criterion,
 179
- Frustration tests, 118
- Gait and personality, 25, 34
- Garment machine operators tests,
 268
- Graphic rating scale, 366 ff.
- Graphology, 439
- Group vs. individual tests, 122
- Guessing test answers, 159
- Guidance, vocational, 8
- Halo effect, 39, 389
- Height and vocational success, 419
- Hierarchy of intelligence, 286
- History of employment psychology,
 52 ff.
- Honesty tests, 114
- Incentive, in industrial work, 174;
 in tests, 151
- Individual vs. group tests, 122
- Industrial school workers' intelli-
 gence, 305
- Information as test of interest, 332,
 342
- Ingenuity, tests of, 91
- Initial success in job predictive of
 later, 415 ff.
- Instructions for tests, 144 ff.; clarity,
 145; form of, 147; oral emphasis,
 127; oral vs. printed, 151
- Insurance salesmen, interest tests,
 340
- Intelligence, alumni groups, 293;
 and criterion, 299; faculty vs.
 students, 291; hierarchy, 286;
 kinds, 99; mechanical, 99; nature
 of, 93; optimum vs. maximum,
 308; and physiognomy, 35; quo-
 tient, 106; school indications of,
 412; social, 100; surveys of in-
 dustries, 298; test validity, 299;
 and turnover, 308
- Intercorrelations of tests, 259 ff.
- Interest, and ability, 322; factor
 analysis, 330; indirect measures
 of, 334, 343; information tests,
 332, 342; inventory, Strong's, 327,
 338; measures, 324 ff.; and oppor-
 tunity, 319; permanence, 320;
 questionnaire, 324; test validity,
 338 ff.
- Interpretation of test scores, 238
- Interview, employment, 450 ff.;
 forms, 457; improvement of, 454
 ff.; morale, 469; unreliability,
 451 ff.
- Interviewer's rating scale, 460
- Item analysis in tests, 277
- Job analysis, 508 ff.; for selecting
 tests, 221 ff.
- Job families, 430
- Justification of ratings, 378
- Laboratory for testing, 227
- Lathe operator test, 235
- Learning, tests of, 83
- Linear scale, criterion estimates, 182
- Literacy of persons taking tests, 212

- Machine scoring of tests, 156
- Maladjustment, vocational, 8
- Man-to-man rating scale, 359 ff.
- Married vs. single employees, 425 ff.
- Master scale, 359
- Mechanical intelligence, 99
- Memory tests, 86
- Mental age, 106
- Mental components of job, tests for, 217, 254
- Messengers, intelligence, 310
- Misleads, selection of, 140
- Motor control tests, 76
- Motormen tests, 234
- Multiple-choice test, 129
- Multiple correlation, 265
- Multiple ratio technique, 266
- Munsterberg, 56

- National Research Council, 64
- Normal frequency curve, 181, 192

- Objective traits, 354, 386
- Observation of subject during tests, 124
- Office boys, intelligence, 301
- Old age, mental efficiency, 211
- Omnibus test, 96, 133, 142
- Optimum vs. maximum intelligence, 308
- Oral tests, 128
- Order of tests, 267
- Orientation of employment psychologist, 220

- Pad for scoring tests, 155
- Partial correlation, 60, 162, 259, 560; factors increasing size of, 264
- Percentile scores, 105
- Performance tests, 130; intelligence, 98
- Permanence of interests, 320
- Persistence tests, 115
- Personal history blank, 418 ff.; weights for items, 431
- Personality tests, 109 ff.; objective vs. subjective, 112; salesmen, 276
- Personnel blank, 418 ff.; academic, 434
- Personnel Research Federation, 65
- Photographs, estimate of traits from, 31

- Phrenology, 21 ff.
- Physiognomy, 17, 24 ff.
- Policemen, intelligence, 311; tests, 273
- Postal employees tests, 274
- Power tests, 137
- Practice, effect on test variability, 150
- Predicting criterion from tests, 240
- Prescored tests, 155
- Pressmen tests, 269
- "Primary abilities," 109
- Probability, of guessing test answers, 160; of success on job, 237
- Production criteria, combination of, 177
- Products-moments correlation, 230
- Pseudo-psychology, 14 ff.
- Psychical research, 20
- Psychological Corporation, 65
- Punch-press operator tests, 235
- Pursuit tests, 78

- Rank difference correlation, 229
- Ranking procedure, difficulty with, 182
- Rapport, 75, 136, 152; with management, 218
- Raters, training, 394
- Rating scale, acquaintance factor, 391; assignment to classes, 365; check list, 375; correction of, 379; defined groups, 362; errors in, 386; graphic, 366 ff.; interviewer's, 460; man-to-man, 359 ff.; for officers, 61; purpose, 347; reliability, 378; selection of traits, 350; time required, 388; validity, 384; weights for traits, 353, 374
- Ratings, justification of, 378
- Reaction time, 87
- Reasoning, tests of, 89
- Recklessness tests, 117
- Recommendation forms, 447
- Recommendations, 444 ff.
- Regression equation, 233, 264, 559
- Reliability, of criterion, 178, 185; of rating scales, 378; split-half, 165; test-retest, 165; of tests, 133, 155, 164 ff., 228; of trade test, 503

- Salesmen, height, 420; intelligence, 303, 312; intelligence hierarchy,

- 294; married, 426; personal history items, 431; rating scale, 385; tests, 275
- Sample, standard of occupations, 342
- Satisfaction and age, 425
- School record for vocational prediction, 406 ff.
- Scoring tests, 153 ff.; ease, 154; key, 154; machine, 156; objective method, 227; pad, 155; speed vs. accuracy, 158; stencil, 154
- Self-control tests, 115
- Sensitivity of test, 144
- Sensory capacity tests, 79 ff.
- Sensory defects of persons taking tests, 212
- Sex differences, interests, 206; subjects for tests, 206
- Shock absorber test, 137, 228
- Significance of difference, 328
- Similarity of test and job, 226
- Social intelligence, 100
- Space perception tests, 89
- Speed tests, 137
- Speed vs. accuracy scores, 158
- Spiral omnibus test, 143
- Spiritualism, 19
- Split-half reliability, 165
- Stability and intelligence, 308
- Standard deviation, computation of, 191; of difference, 329
- Standard error of estimate, 238, 247
- Standard sample of occupations, 342
- Standard scores, 190, 246
- Stature, 419
- Stencil for scoring tests, 154
- Stereotyping of output, 174
- Strong's Interest Inventory, 327, 338
- Subjects for evaluating tests, 200 ff.
- Substation operators tests, 268
- Successive hurdles method, 267
- Supervisor's estimates as criterion, 179
- Team work, influencing criterion, 176
- Telegraphers tests, 274
- Telepathy, 20
- Terminology of tests, justification of, 74
- Tests, administration, (22 ff.; early development of, 54; group vs. individual, 122; individual vs. group, 122; instructions, 144 ff.; interpretation of scores, 238; material, 137 ff.; oral, 128; order of, 267; performance, 130; power, 137; practice effect, 149; reliability, 133, 155, 164 ff.; scoring, 153 ff.; selection of, 255; sensitivity, 144; similarity to job, 226; speed, 137; standard conditions, 135; trade, 60, 477 ff.
- Time-limit tests, 132; vs. work-limit tests, validity, 140
- Tire factory, intelligence tests, 307
- Tire finishers tests, 255
- Total mental situation, tests for, 216 ff.
- Trade tests, 60, 477 ff.; criterion for, 484; reliability, 503; validity, 504
- Training raters, 394
- Traits, definition of, 356; objective, 354, 386; selection for rating scales, 350; weights in rating scale, 353, 374
- Turnover, age, 424; intelligence, 308
- Unique tests, 9
- United States Employment Service, 66
- Unreliability of tests, causes of, 166
- Upper critical score, 313
- Validity, rating scale, 384; tests, 5; time- vs. work-limit tests, 140; trade test, 504
- Vocational guidance vs. selection, 8
- Waitresses, intelligence, 312
- Weighting speed and accuracy, 160
- Weighting tests, 259 ff.
- Weights of criterion variables, 196
- Welfare, industrial psychology and, 7 ff.
- White-collar interest, 10, 320
- Work limit tests, 132; determination of limits, 134; reliability, 133
- Work sample as criterion, 177
- Work- vs. time-limit tests, validity, 140
- X-ray operator tests, 236

